



Metabolic influences on ageing in *Caenorhabditis elegans*

A time series metabolomics and metabolic modelling study



Janna Hastings

Supervisors: Dr Olivia Casanueva
Dr Nicolas le Novère

Epigenetics Department
Babraham Institute

This dissertation is submitted for the degree of
Doctor of Philosophy

This thesis is dedicated to Sven, Aidan, Aliya and Yann
A happy family is the fertile soil for the flowering of the mind

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

It is not substantially the same as any that I have submitted, or is being concurrently submitted, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. Moreover, no substantial part of my thesis has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

In keeping with the requirements of the Degree Committee in Biology, this thesis contains fewer than 60,000 words excluding appendices, bibliography, footnotes, tables and equations.

Janna Hastings
March 2019

Metabolic influences on ageing in *Caenorhabditis elegans*

Janna Hastings

Abstract

Ageing presents one of the most fundamental public health challenges of our time. Progress in living standards, combating infectious disease and promoting safety, and general nutritional availability, has led to an increase in lifespan across the developed world. However, this has been accompanied by an increase in the duration of late-life frailty and associated conditions affecting health. Metabolism is known to be a key mediator of ageing across the diversity of living species. Many of the pathways that extend lifespan and promote healthspan are known to be metabolic and relate to managing the balance of energy availability to optimise resource usage and survival during times of scarcity.

The model organism *Caenorhabditis elegans*, a small transparent nematode worm that ordinarily lives in the soil and eats bacteria, is one of the most common organisms used in the study of ageing as it is easy to culture in laboratory conditions and has a short lifespan of around three weeks under normal conditions. In this thesis, I analyse in detail the metabolic changes that occur during ageing in *C. elegans*, using a multi-omics metabolomics and transcriptomics time series of measurements in three *C. elegans* strains, and mathematical modelling.

Whole-genome metabolic models are representations of all the metabolic reactions taking place within an organism together with their metabolic inputs and outputs, and enzymatic catalysts. I describe the development and validation of a community-wide shared whole-genome metabolic model for *C. elegans*. Using this model together with measured gene expression levels for each enzyme that catalyses a reaction, it is possible to predict intracellular reaction fluxes using a method called Flux Balance Analysis (FBA). I describe a novel method for the integration of metabolomics data with FBA, and the results of a comparative analysis of the resulting fluxes in normal wild-type ageing. I then go on to describe the differences to a germline-free strain that is long-lived and metabolically different.

Finally, I have used the model to probe the metabolic flexibility and evidence for trans-omics bidirectional regulation between the transcriptomic and metabolomic layers.

Acknowledgements

I am enormously grateful to my supervisors for their support and guidance. Thanks to Olivia Casanueva for welcoming a purely computational student to her lab, allowing me a privileged insider's view into the processes and practices of an experimental research group in biology. Thanks to Nicolas le Novère for sharing his love of, and multifaceted knowledge in, all aspects of bioinformatics and mathematical modelling. A very special thank you to Erica Watson for standing in as my University Supervisor on short notice.

I am indebted to the dedicated team of scientists who laboured long hours in the laboratory to grow the samples and produce the data on which this thesis is based. In particular I would like to thank Abraham Mains for his enormous efforts in producing a carefully staged, detailed time course of measurements of gene expression and metabolite concentrations. Thanks also to Bhupinder Virk, Sharlene Murdoch and Laetitia Chauve for sharing their biological expertise and performing experiments; to them as well as Manusnan Suriyalaksh, thanks for sharing specific figures, as specified in the text.

For the development of the consensus whole-genome metabolic model, I give my thanks to all members of the WormJam consortium who contributed to the many discussions and hands-on community curation effort. In particular, I would like to thank Juliane Gebauer and Christoph Kaleta for sharing the ElegCyc model, Chintan Joshi and Nathan Lewis for sharing their merged model, Michael Witting for his painstaking and detailed curation efforts and general enthusiasm, and Jake Hattwell for many aspects of model validation and general problem-solving. I would also like to thank Nicolas Rodriguez for creating automated tools to support the curation and validation of the model, and Juliette Pearce for sharing her computational approaches to differential flux analysis, pathway enrichment, and flux visualisation.

To all the members of the Casanueva and le Novère groups at the Babraham Institute, my wholehearted thanks for the positive working environment, scientific knowledge, and many informative and interesting discussions over the years. To my fellow students at Darwin College, especially the members of the 2017-2018 and 2018-2019 DCSA executive committees, thanks for the fun times that helped to keep me sane.

None of this would have been possible without the ongoing love and support of my partner, Sven, and the forbearance of my three children, Aidan, Aliya and Yann.

Table of Contents

| | |
|--|------------|
| List of Figures | xv |
| List of Tables | xix |
| List of Abbreviations | xxi |
| 1 Introduction | 1 |
| 1.1 <i>C. elegans</i> is a popular model organism for ageing research | 1 |
| 1.2 Metabolism is a key mediator of longevity | 3 |
| 1.2.1 Dietary restriction and nutrient sensing pathways | 5 |
| 1.2.2 Germline signalling and lipid metabolism | 7 |
| 1.2.3 Mitochondrial function | 9 |
| 1.2.4 Epigenetic modifiers are metabolic sensors | 10 |
| 1.2.5 Summary: metabolism and ageing in <i>C. elegans</i> | 11 |
| 1.3 Systems biology approaches to the study of metabolism | 12 |
| 1.3.1 Metabolomics as a tool to study metabolic changes | 14 |
| 1.3.2 Whole-genome metabolic modelling as an integrative systems biology approach | 15 |
| 1.4 About this thesis: Metabolic modelling with multi-omics data in ageing <i>C. elegans</i> | 16 |
| 1.4.1 Objectives | 16 |
| 1.4.2 Publications | 17 |
| 1.4.3 Organisation of the thesis | 17 |
| 2 Metabolic changes during ageing | 19 |
| 2.1 Introduction | 19 |
| 2.2 Methods | 21 |

| | | |
|----------|---|-----------|
| 2.2.1 | Strains used in the study | 21 |
| 2.2.2 | Worm maintenance and sample information | 22 |
| 2.2.3 | Metabolomics assay | 23 |
| 2.2.4 | Data pretreatment and quality control | 24 |
| 2.2.5 | Multivariate analysis methods: PCA and PLS-DA | 26 |
| 2.2.6 | Literature and database search for longevity-modulating metabolites | 28 |
| 2.2.7 | Pathway enrichment – Fisher’s exact test | 28 |
| 2.2.8 | Determining metabolites with oscillations within days in early ageing | 28 |
| 2.3 | Results and Discussion | 29 |
| 2.3.1 | Age is the primary driver of variability in the metabolome | 29 |
| 2.3.2 | 43 metabolites are significantly changing with age | 33 |
| 2.3.3 | Polar amino acids change the most with age | 35 |
| 2.3.4 | Age-associated metabolites are enriched for longevity modulators | 37 |
| 2.3.5 | Temporal patterns over the full course of ageing | 39 |
| 2.3.6 | Strain differences and patterns during early ageing | 42 |
| 2.3.7 | Some metabolites vary with time of day, in what may reflect circadian metabolic oscillations | 47 |
| 2.4 | Conclusion | 51 |
| 3 | A consensus genome-scale model of <i>C. elegans</i> metabolism | 55 |
| 3.1 | Introduction | 55 |
| 3.2 | Methods | 56 |
| 3.2.1 | Whole-genome metabolic models | 56 |
| 3.2.2 | The WormJam model annotation and development pipeline | 57 |
| 3.2.3 | Model validation | 58 |
| 3.2.4 | Visualisation of arbitrary reaction groups | 59 |
| 3.3 | Results and Discussion | 61 |
| 3.3.1 | Initiation of the WormJam community | 61 |
| 3.3.2 | Overlaps and differences between published models | 62 |
| 3.3.3 | Manual annotation of the model | 64 |
| 3.3.4 | The WormJam consensus reconstruction: current status | 69 |
| 3.3.5 | Model validation | 71 |
| 3.4 | Conclusion | 77 |

| | | |
|----------|---|------------|
| 4 | Metabolic flux changes with ageing | 79 |
| 4.1 | Introduction | 79 |
| 4.2 | Methods | 79 |
| 4.2.1 | Flux Balance Analysis | 79 |
| 4.2.2 | Integration of -omics data with FBA | 83 |
| 4.2.3 | Differential flux pathway analysis | 85 |
| 4.2.4 | Flux visualisation | 87 |
| 4.3 | Results and Discussion | 87 |
| 4.3.1 | A novel method for integrating metabolomics data with FBA | 87 |
| 4.3.2 | Overall growth rate and environmental exchanges (Standard FBA) | 97 |
| 4.3.3 | Age-related changes in pathway fluxes (Metab FBA) | 99 |
| 4.4 | Conclusion | 105 |
| 5 | Metabolic regulation of ageing | 109 |
| 5.1 | Introduction | 109 |
| 5.2 | Methods | 109 |
| 5.2.1 | Regulation and Bottlenecks in Flux Predictions | 109 |
| 5.2.2 | Clustering and network inference | 111 |
| 5.3 | Results and Discussion | 111 |
| 5.3.1 | Flexibility and constraint in the metabolic network | 111 |
| 5.3.2 | Prediction of novel longevity modulators based on metabolite network distances | 116 |
| 5.3.3 | Exploring mechanisms of metabolic regulation by predicting gene–metabolite interactions | 120 |
| 5.4 | Conclusion | 122 |
| 6 | General Discussion | 123 |
| | References | 127 |
| | Appendix A Supplementary Tables And Figures | 141 |

List of Figures

| | | |
|------|--|----|
| 1.1 | <i>C. elegans</i> anatomy, development and adult life course. | 2 |
| 1.2 | The insulin/IGF signalling pathway regulates development and ageing. . . . | 4 |
| 1.3 | Dietary restriction and nutrient sensing pathways. | 6 |
| 1.4 | Summary of metabolic pathways impacting lifespan and healthspan. | 11 |
| 2.1 | FEM and GEM strains have lifespans comparable to wild type | 22 |
| 2.2 | Counts of missing values per metabolite and sample. | 25 |
| 2.3 | Distances to nearest replicate within a sample group. | 26 |
| 2.4 | Correlation analysis of the metabolomics dataset | 30 |
| 2.5 | Principal component analysis of the metabolomics dataset | 31 |
| 2.6 | Fold changes comparing young to old concentrations for age-associated metabolites. | 34 |
| 2.7 | Pathway enrichment for age-associated metabolites. | 36 |
| 2.8 | Bubble plot for pathway enrichment. | 38 |
| 2.9 | Metabolite intensities clustered by hour of collection. | 40 |
| 2.10 | <i>C. elegans</i> body size at days 1, 5 and 10. | 41 |
| 2.11 | Heat map of metabolites that differ by strain | 44 |
| 2.12 | Selected metabolites showing characteristic between-strain differences. . . . | 45 |
| 2.13 | Availability of energy metabolites. | 46 |
| 2.14 | Time of day effects in early ageing metabolism. | 48 |
| 2.15 | <i>xbp-1</i> as one candidate transcriptional mediator of time of day variation in metabolite concentrations. | 52 |
| 3.1 | A schematic illustrating the structure of whole-genome metabolic models. . | 56 |
| 3.2 | A subset of pathways in the WormJam model, visualised with the MetDraw tool. | 60 |

| | | |
|------|---|-----|
| 3.3 | Overlaps between two models based on name matching. | 63 |
| 3.4 | Overview of the infrastructure for manual model curation. | 65 |
| 3.5 | Counts of identifiers for metabolites in WormJam model | 66 |
| 3.6 | Visualisation of arbitrary groups of reaction chains. | 68 |
| 3.7 | Overview of the entire WormJam consensus model. | 69 |
| 3.8 | Sizes of compartments and pathways in WormJam model. | 70 |
| 3.9 | Growth of the model as a function of allowed inputs. | 72 |
| 3.10 | Oxidative ATP synthesis and oxygen consumption | 73 |
| 3.11 | Percentages of blocked reactions in model pathways. | 74 |
| 4.1 | A schematic illustrating Flux Balance Analysis. | 80 |
| 4.2 | Integration of -omics data with Flux Balance Analysis. | 83 |
| 4.3 | Pathway enrichment comparing fluxes between days 5 and 10 in standard FBA | 88 |
| 4.4 | TCA cycle standard FBA fluxes compared between FEM days 5 and 10 . . . | 89 |
| 4.5 | Overview of Metab FBA approach | 90 |
| 4.6 | Comparison of overall indicators for Standard and Metab FBA methods . . . | 92 |
| 4.7 | Flux enrichment for Metab FBA fluxes, FEM days 5 to 10 | 94 |
| 4.8 | Predicted overall fluxes for three pathways | 94 |
| 4.9 | TCA cycle fluxes with three different objective functions | 95 |
| 4.10 | Overall metrics for the growth of the organism | 98 |
| 4.11 | Fluxes through key reactions indicating energy usage | 100 |
| 4.12 | Pathways with flux differences between consecutive time points in FEM . . . | 101 |
| 4.13 | Metab FBA total sum of fluxes predicted for selected pathways | 102 |
| 4.14 | Pathways with flux differences between consecutive time points in GLP . . . | 104 |
| 4.15 | Predicted reaction network usage and total flux sum | 105 |
| 5.1 | Flux Variability range sums per pathway | 113 |
| 5.2 | Pathway enrichment for reactions with reduced costs below zero | 115 |
| 5.3 | Network of correlations and anti-correlations for all metabolites with correla- tion threshold > 0.65 and p(adj) of corr < 0.05. | 117 |
| 5.4 | Network-based predictions of longevity modulating scores for metabolites . | 118 |
| 5.5 | Validation of predicted longevity modulating scores for metabolites | 119 |
| 5.6 | Network of correlations between genes and metabolites with correlation threshold > 0.7 and p(adj) of corr < 0.05. | 120 |

| | | |
|------|---|-----|
| A.1 | Additional PCA plots for the third and fourth PC | 159 |
| A.2 | The PLS-DA model of the ageing metabolome with hour of collection as response | 160 |
| A.3 | The PLS-DA model of the early ageing metabolome with strain as response . | 161 |
| A.4 | Individual level plots of metabolites that differ by strain | 162 |
| A.5 | Clustered heat map of full early ageing dataset | 163 |
| A.6 | PCA of flux values predicted for individual replicate samples. | 164 |
| A.7 | Comparison of TCA cycle in aged worms to <i>idh-1</i> mutants. | 165 |
| A.8 | Comparison of standard FBA fluxes for several pathways between FEM and GLP at 49 h | 166 |
| A.9 | Comparison of Metab FBA fluxes for several pathways between GLP 49 and 65 h | 167 |
| A.10 | Numbers of reactions with reduced costs < 0 per strain and time point . . . | 168 |

List of Tables

| | | |
|-----|--|-----|
| 2.1 | Ageing metabolomics studies in <i>C. elegans</i> | 19 |
| 2.2 | Number of replicates per strain and time point for metabolomics dataset. . . | 29 |
| 2.3 | Numbers of metabolites in age-associated and longevity modulating groups. | 37 |
| 3.1 | Metrics for the size of two published models | 62 |
| 3.2 | Metrics for the size of the WormJam consensus reconstruction | 70 |
| 5.1 | Centrality scores for metabolites in gene-metabolite network. | 121 |
| A.1 | Names, pathways and additional information for measured metabolites in metabolomics assay | 141 |
| A.2 | Significantly age-associated metabolites relationship to known metabolic changes in the literature . ↑-metabolite level increases with advanced age. ↓-metabolite level decreases with advanced age. | 145 |
| A.3 | Longevity modulating metabolites in the literature and DrugAge database. | 147 |
| A.4 | Pathway metrics | 149 |
| A.5 | Essential genes in WormJam model | 152 |
| A.6 | Mapping of metabolomics measured metabolites to WormJam model metabo- lites | 154 |
| A.7 | Predicted network-based longevity modulation scores for metabolites | 157 |

List of Abbreviations

C. elegans gene names

| | |
|----------------|--|
| <i>aak-2</i> | Active subunit of AMP-activated kinase (AMPK) |
| <i>age-1</i> | Phosphatidylinositol 3-kinase |
| <i>akt-1,2</i> | Serine/threonine kinase Akt/PKB |
| <i>crtc-1</i> | CREB-regulated transcriptional coactivator |
| <i>daf-15</i> | TORC1-binding partner Raptor |
| <i>daf-16</i> | Transcription factor FOXO |
| <i>daf-2</i> | Insulin/IGF-1 receptor |
| <i>fat-6,7</i> | Stearoyl-CoA- δ 9-desaturase |
| <i>fem-3</i> | Masculinisation of germline |
| <i>gem-1</i> | Solute carrier |
| <i>glp-1</i> | Notch signalling receptor |
| <i>gon-2</i> | Cation channel required for development of gonad |
| <i>hsf-1</i> | Heat shock factor 1 |
| <i>lipl-4</i> | Lipase homologous to human lysosomal acid lipase |
| <i>pha-4</i> | Transcription factor FOXA |
| <i>rict-1</i> | TORC2-binding partner Rictor |
| <i>skn-1</i> | Transcription factor Nrf |
| <i>xbp-1</i> | X-box Binding Protein homolog |

Acronyms/Abbreviations

| | |
|------|-------------------------|
| 3HBA | 3-hydroxybutyric acid |
| ADP | Adenosine diphosphate |
| AMP | Adenosine monophosphate |
| AMPK | AMP-activated kinase |
| AOV | Analysis of variance |
| ATP | Adenosine triphosphate |
| DR | Dietary restriction |
| ER | Endoplasmic reticulum |

| | |
|-------------------|---|
| FBA | Flux Balance Analysis |
| FEM | Strain used in our study, <i>fem-3(q20)ts</i> |
| FEM | Strain used in our study, <i>glp-1(e2144)ts</i> |
| FuDR | Fluorodeoxyuridine |
| FVA | Flux Variability Analysis |
| FVA | Flux Variability Analysis |
| FV | Flux Variability |
| GEM | Strain used in our study, <i>gon-2(q388)ts;gem-1(bc364)</i> |
| GSMM | Genome-Scale Metabolic Model |
| IGF | Insulin growth factor |
| IIS | Insulin/IGF-1 signalling |
| MS | Mass Spectrometry |
| MUFA | Mono-unsaturated fatty acid |
| NAD ⁺ | Nicotinamide adenine dinucleotide, oxidized |
| NADH | Nicotinamide adenine dinucleotide, reduced |
| NADP ⁺ | Nicotinamide adenine dinucleotide phosphate, oxidized |
| NADPH | Nicotinamide adenine dinucleotide phosphate, reduced |
| NMR | Nuclear Magnetic Resonance |
| PCA | Principal Components Analysis |
| PI3 | Phosphoinositide 3 |
| PLS-DA | Partial Least Squares – Discriminant Analysis |
| PUFA | Poly-unsaturated fatty acid |
| ROS | Reactive oxygen species |
| SAM | S-adenosyl-methionine |
| SBML | Systems Biology Markup Language |
| TCA | Tricarboxylic acid |
| TOR | Target of rapamycin |
| UDP | Uridine diphosphate |
| VIP | Variable Importance in Projection |
| WT | Wild type |
| YA | Young adult |

1. Introduction

Across the developed world, we are witnessing a significant increase in the burden of age-associated diseases (Beard and Bloom, 2015). Average life expectancies have increased, thanks to advances in medical care and living standards. However, the period of relatively good health ('healthspan'; Hansen and Kennedy, 2016) has not been extended concordantly with that of life expectancy, leading to a public health crisis (Partridge et al., 2018).

Ageing across multiple species is accompanied by frailty, decline in normal physiological functioning, and disease (López-Otín et al., 2013). The diseases that occur disproportionately in aged human populations include cardiovascular disease, metabolic syndrome and diabetes, cancer and neurodegeneration (Hodes et al., 2016). Yet, there is ample evidence that lifestyle and metabolic factors may delay the onset of age-associated disease and prolong health (Chedraui and Pérez-López, 2013; Dato et al., 2013; Quach et al., 2017; Rea, 2017).

Ageing comprises a complex, interconnected series of dysfunctions that develop as organisms grow older, and while individual aspects of this process are being elucidated in great detail, the field still lacks a comprehensive systems-level integrated understanding (Cohen, 2016; Kirkwood, 2008, 2011). It is becoming evident that the mysteries of normal and perturbed ageing need to be studied as a complex temporal sequence of events to tease apart the interrelated regulatory processes (Hastings et al., 2019b). Furthering this objective to extend knowledge of the basic biology of ageing, and the ways in which genetic and metabolic factors interact in a time resolved, whole-systems level to delay ageing-associated functional decline, promises to yield improvements in health over the full lifespan (Chauhan et al., 2015; Niccoli and Partridge, 2012; Zierer et al., 2015).

1.1 *C. elegans* is a popular model organism for ageing research

Much of what is known about the biology of ageing has been learned from studies using the model organism *Caenorhabditis elegans* (*C. elegans*), a transparent nematode worm about 1mm in length, which has become the première model organism used in ageing research (Braeckman and Vanfleteren, 2007; Gruber et al., 2015; Johnson, 2013; Torgovnick et al., 2013). *C. elegans* is a popular metazoan model organism to study ageing as it bears sufficient molecular resemblance to mammalian species in many aspects, while being amenable to experimentation due to its relatively simple body plan and much shorter lifespan.

C. elegans has an average natural lifespan of only 2-3 weeks, making comparative lifespan studies feasible in relatively short timeframes. It is easily grown in laboratory

cultures, and can be frozen for long time periods in liquid nitrogen, hence making the maintenance of isogenic populations straightforward (Corsi et al., 2015). It was one of the first organisms to be genetically sequenced, and there are a wide variety of genetic tools available for manipulation. Many of the important markers of ageing and senescence in humans are also present in *C. elegans*, including decreased mobility, frailty, and increased susceptibility to infection (Collins et al., 2008). Moreover, age-related pathology can be directly screened in *C. elegans* due to the transparency of the worm body. This transparency supports the widespread use of fluorescent reporters and the development of automated imaging approaches for high-throughput phenotyping (Stroustrup et al., 2013). *C. elegans* is robustly responsive to multiple interventions extending lifespan, and at the molecular level, many of the pathways that extend lifespan in *C. elegans* are conserved across species and have similar effects in mammals (Tissenbaum and Guarente, 2002; Torgovnick et al., 2013).

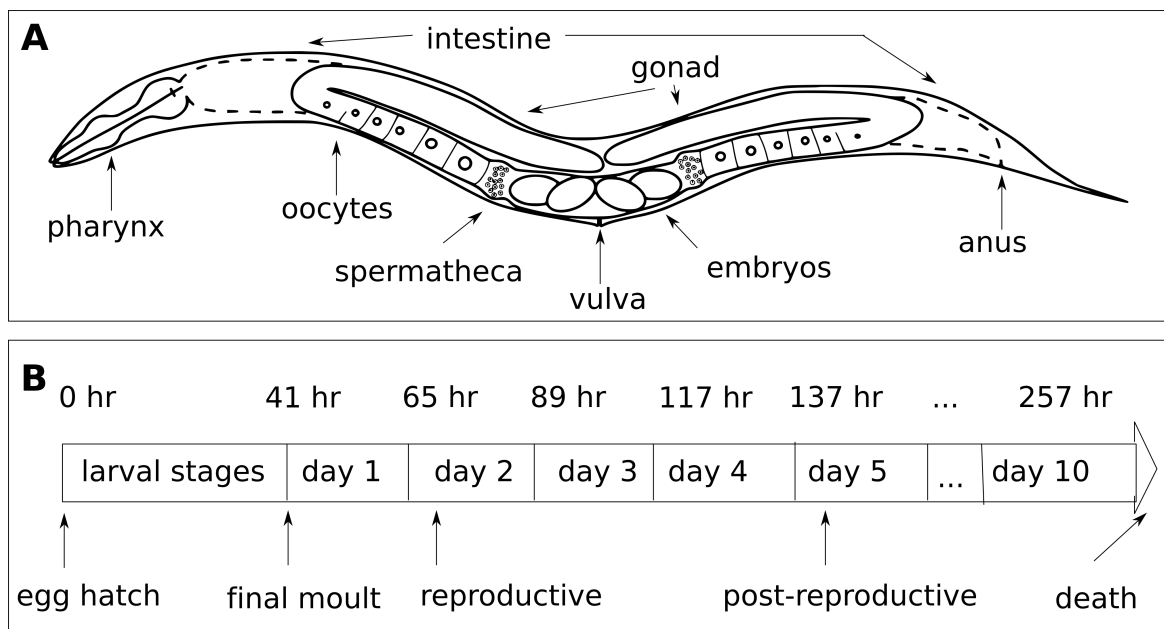


Figure 1.1 *C. elegans* anatomy, development and adult life course.

(A) The basic anatomy of adult hermaphrodite *C. elegans*. The pharynx is where bacteria are ground up to be ready for digestion in the intestine, which extends throughout the body. The somatic gonad produces both sperm and oocytes which are then self-fertilised to generate embryos. (B) The developmental and adult life course timings are illustrated from the time of egg hatching into the first larval stage. The adult worms are reproductive between days 2 and 4 of adulthood. Loss of functionality due to ageing ensues a few days later, with onset of age-related mortality post-day-10. Approximate timings are indicated in hours.

The majority of *C. elegans* individuals are self-fertile hermaphrodites. In wild type (WT) conditions, males make up 0.1-0.2% of the population (Corsi et al., 2015). *C. elegans* are ectotherms, and the durations of both developmental stages and lifespan are influenced by the surrounding environmental temperature (Klass, 1977; Zhang et al., 2015), with development proceeding more rapidly at warmer temperatures. Figure 1.1A shows the

anatomy of WT adult hermaphrodite *C. elegans*, and Figure 1.1B shows an indicative WT timeline of development, adulthood, reproduction and age-associated decline.

Just as is the case for humans, a complex and interconnected variety of organismal pathologies characterise normal ageing in *C. elegans*: the pharynx, through which *C. elegans* consume their bacterial diet, degenerates, while the intestine and gonad atrophy, and tumours develop in the uterus (Ezcurra et al., 2018). Alongside atrophy of the intestine, fatty pools of lipoproteins accumulate in the body cavity. Cellular pathologies are also evident: stochastic molecular damage accumulates, and mitochondria become fragmented and lose volume (Regmi et al., 2014; Yasuda et al., 2006) with a concurrent loss of electron chain oxygen consumption and ATP production (Braeckman et al., 2002; Houthoofd et al., 2005). At the same time, locomotion declines alongside a loss in muscle tissue integrity (Glenn et al., 2004; Hsu et al., 2009). These changes do not merely occur towards the end of the normal worm lifespan – in fact, many are fully evident prior to median lifespan, and appear in some cases to anticipate, rather than result from, changes in messenger RNA and protein levels that occur with advancing age (Ezcurra et al., 2018), hinting that the age-related pathologies may be causes rather than consequences of deregulated transcriptional processes. And it is not only the changes that occur late in ageing when morphological decline is evident that are relevant, but also the changes during early ageing. For example, there is a very early collapse in the ability of the organism to respond to stress, by up-regulating the transcription of stress response proteins, that takes place roughly at the same time as the onset of reproduction (day two of adulthood), yet can be delayed by interventions that extend lifespan (Ben-Zvi et al., 2009).

C. elegans has consistently been at the forefront of ageing research, as the very first genetic mutation that results in a longer lifespan was discovered in this organism. This long-lived mutation was a loss of function in the *age-1* gene, which encodes for a PI3 kinase (Friedman and Johnson, 1988; Johnson, 2013). This gene encodes for a key component in the signalling cascade that listens for and responds to the presence of insulin, an indicator of the availability of nutrients, and a fundamental regulator of diverse metabolic pathways. With this discovery, ageing was revealed as not inevitable in its decline, but rather as a genetically controlled, plastic phenomenon amenable to possible interventions (Kenyon, 2011). Shortly thereafter, the subsequent discovery of a second longevity modulator further cemented the relevance of metabolism for ageing.

1.2 Metabolism is a key mediator of longevity

After *age-1*, the next longevity enhancing mutant to be discovered was *daf-2*, which encodes the *C. elegans* insulin/IGF receptor ortholog (Kenyon, 2011; Kenyon et al., 1993).

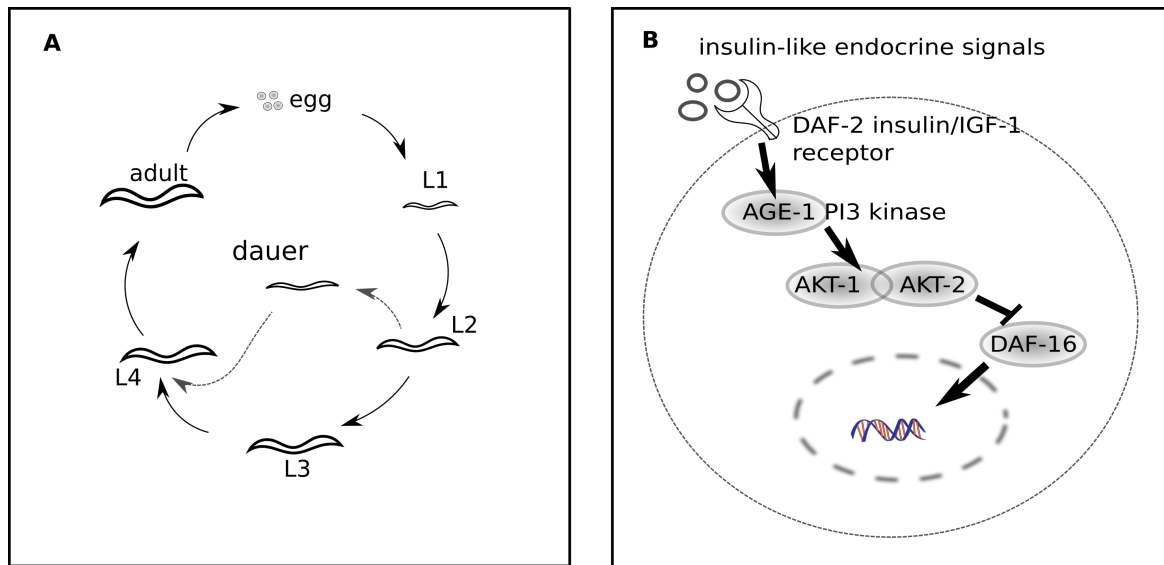


Figure 1.2 The insulin/IGF signalling pathway regulates development and ageing.

(A) The life cycle of *C. elegans* typically proceeds from hatching through four larval stages to adulthood and the onset of reproduction. The dauer stage is an alternative larval stage entered in conditions of food scarcity. In this stage, the organism is hypometabolic and can survive for up to four months, many times its normal adult lifespan. When food becomes available again, the organism exits into the L4 larval stage and re-enters normal development. Entry into the dauer stage is controlled by the IIS pathway. (B) In the IIS pathway, signals about food abundance are received by the DAF-2 receptor, which activates the AGE-1 PI3 kinase signalling cascade that acts through AKT-1 and AKT-2 to inhibit translocation of the transcription factor DAF-16 to the nucleus. When signalling in this pathway is reduced, DAF-16 enters the nucleus and activates transcription of stress response and pro-longevity genes.

During development, if nutrients are scarce, *C. elegans* may enter an alternative developmental stage called 'dauer'. Animals can stay in the dauer stage for up to several months, surviving from their stored lipid reserves with a very low rate of metabolism (Figure 1.2A). If food becomes abundant in the environment again, they will exit the dauer stage and continue with normal development.

The gene *daf-2* is one of several genes known as constitutive for dauer formation, as loss of function mutations in these genes result in dauer formation during development even when nutrients are abundant. Using a temperature-sensitive mutation in this gene to enable selective activation in adulthood, Kenyon et al. (1993) and others were able to show that this mutation could extend lifespan in adults, indicating that nutrient sensing pathways could fundamentally regulate not only development, but also the process of ageing.

It was subsequently determined that *daf-2* is a part of the insulin/IGF-1 signalling (IIS) pathway (Figure 1.2B), a widely conserved nutrient sensing pathway that also includes *age-1* (Kaletsky and Murphy, 2010; Panowski and Dillin, 2009). *daf-2* acts upstream of *age-1*, while downstream of *age-1* is the transcription factor *daf-16* that has been found to

have prolific regulatory effects on stress resistance, cellular metabolism and proteostasis (the maintenance of proteins in their correct folded form and the removal of damaged proteins). In humans, insulin is secreted primarily by the pancreas, and insulin signalling has been found to regulate not only glucose homeostasis but also pleiotropically many other pathways including growth and reproduction (Mathew et al., 2017), and, perhaps most surprisingly, cognitive function (Tumminia et al., 2018). Impairments in insulin signalling lead commonly to body weight deregulation and diabetes (Brown et al., 2017), and, in some rare severe genetic mutations, developmental defects which lead to early death (Bathi et al., 2010). Mutations that cause impairments in the insulin growth factor (IGF-1) receptor in humans result in severe reductions in growth, and defects in metabolism (Savage, 2013).

In addition to the IIS pathway, longevity is now known to be influenced by a multifactorial array of heritable and environmental influences interrelated with metabolic factors (Finkel, 2015). The morphological changes mentioned above that are evident in post-reproductive wild-type *C. elegans* also point towards the loss of key metabolic capabilities during normal ageing, e.g. atrophy of the intestine (Ezcurra et al., 2018) and loss of mitochondrial integrity (Yasuda et al., 2006).

The remainder of this section provides a review of the interrelationships between metabolism and ageing.

1.2.1 Dietary restriction and nutrient sensing pathways

One of the earliest discovered and most reliably conserved interventions to extend lifespan is dietary restriction (DR), the reduction of food intake without malnutrition (Omodei and Fontana, 2011). DR reduces the incidence of many age-associated diseases, including cardiovascular disease, cancer and metabolic disease (Colman et al., 2009). Conversely, over-nutrition and obesity lead to complications that shorten life span and exacerbate age-associated decline, such as in cognitive performance (Beilharz et al., 2015).

DR acts to promote longevity through a combination of mechanisms and pathways. It shifts the balance of active metabolism from carbohydrates to lipids, improves metabolic flexibility and insulin sensitivity, inhibits inflammation and oxidative stress, and maintains youthful patterns of acetylation (sirtuin activity), NAD⁺ levels, and autophagy (Fontana et al., 2010; Omodei and Fontana, 2011). During the normal course of ageing, there is a general reprogramming of metabolic function, including a decrease in the efficiency of energy generation, which is believed to be one of the factors that is offset by DR (Feng et al., 2016; Houtkooper et al., 2013). However, disentangling the specific effects of these various pro-longevity pathways to understand their interrelationships and individual contributions to overall longevity phenotypes is an ongoing challenge in the field of ageing research.

In *C. elegans*, early research established that, as with other model organisms, DR robustly extended lifespan (Klass, 1977). Since then, several distinct models for DR have been developed in *C. elegans*, including dietary interventions such as diluting the available bacterial food source or providing an axenic food source only, and genetic mutations such as in the *eat-2* gene that slows the pharyngeal pumping rate (Greer and Brunet, 2009). DR results in decreased activity in nutrient sensing pathways, due to the decreased availability of nutrients. One such pathway is the IIS pathway, already introduced above as that in which the first longevity-promoting genetic interventions were discovered in *C. elegans*. Other nutrient sensing pathways include the TOR (target of rapamycin) signalling pathway and AMPK (AMP-activated kinase).

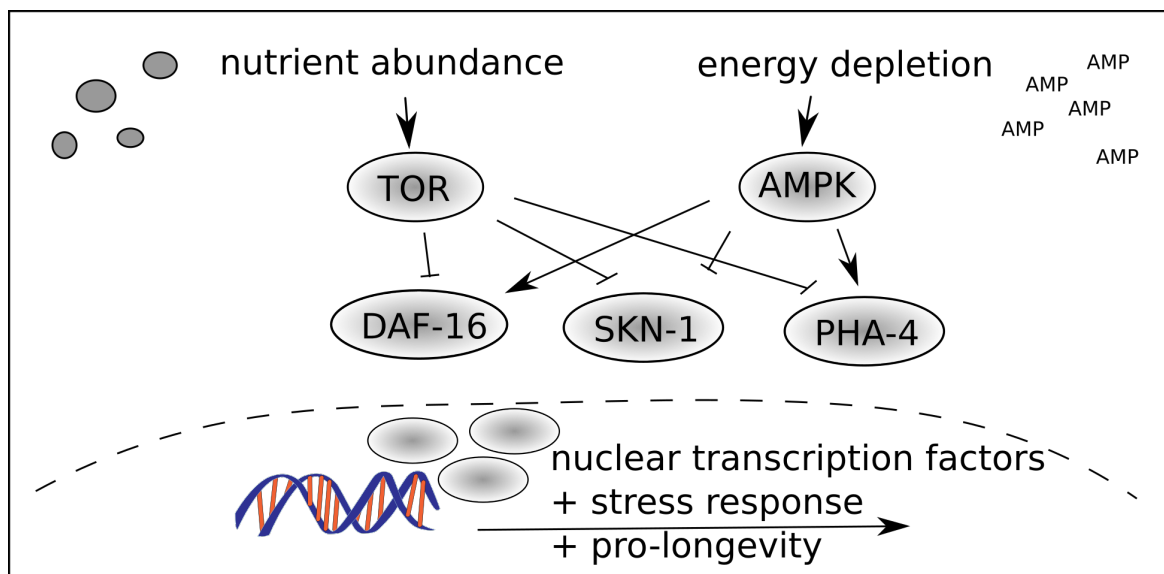


Figure 1.3 Dietary restriction and nutrient sensing pathways.

The TOR pathway, AMPK, and dietary restriction act via convergent interactions to detect nutrient abundance and direct cellular metabolism accordingly, harnessing shared central metabolic and stress response regulators such as *daf-16* and *pha-4*.

The TOR pathway is activated by the availability of nutrients, in particular amino acids and growth factors. The *C. elegans* ortholog of the mammalian TOR gene is *let-363*, and reducing activity of this gene by RNAi in adult worms extends lifespan, but the effect is not increased by simultaneous DR, indicating that it is one of the mechanisms of DR life extension (Vellai et al., 2003). TOR exists in two complexes, TORC1 and TORC2, which have different co-activators – *daf-15* (Raptor) and *rict-1* (Rictor). In addition to longevity, the TOR pathway regulates development, lipid storage, mRNA translation, and autophagy, operating through the downstream regulators *daf-16* (in common with IIS), *pha-4* and *skn-1* (Lapierre and Hansen, 2012).

The energy-sensing enzyme AMP-activated kinase (AMPK) also plays a role in longevity and interacts with the targets of the TOR and IIS pathways. In *C. elegans*, the gene *aak-2*

encodes for a catalytic subunit of AMPK, and its overexpression extends lifespan (Apfeld et al., 2004). AMPK is allosterically inhibited by ATP and activated by AMP, thus acts as a sensor of a high AMP to ATP ratio, which indicates low availability of energy. It acts in part via transcriptional co-regulator *crtc-1* (Mair et al., 2011). It is a widely conserved energy regulation and stress response signalling pathway (Carling, 2017; Garcia and Shaw, 2017) and is frequently aberrant in cancer (Monteverde et al., 2015). AMPK also centrally regulates neuromuscular biology, driving the physiological plasticity in response to exercise, and is thus a target for treatments for neuromuscular conditions such as muscular dystrophy (Dial et al., 2018).

The relationship between nutrient sensing pathways and lifespan is of paramount importance in the quest for pharmacological agents that can be used to slow ageing and prolong health. DR mimetics are drugs that act to inhibit the signalling pathways that indicate nutrient abundance, thereby artificially causing cells to act as though nutrients were scarce (Ingram and Roth, 2015; Ingram et al., 2006). One of the first such DR mimetics to be discovered was rapamycin, a TOR inhibitor which gave its name to the TOR protein complex (Ingram et al., 2006). Other TOR inhibitors include valproic acid (Evason et al., 2008) and α -ketoglutarate (Chin et al., 2014). Other molecules have been found to target other components of the overall nutrient-sensing pathway, for example 2-deoxyglucose (which inhibits glycolysis; Ingram and Roth, 2011), metformin (which enhances insulin action via AMPK; Onken and Driscoll, 2010), and resveratrol (which targets sirtuin pathways; Chung et al., 2012), have all been shown to extend lifespan and lower the risk for age-related diseases.

1.2.2 Germline signalling and lipid metabolism

In *C. elegans*, the removal of proliferating germline stem cells by laser ablation was found to significantly extend lifespan (Hsin and Kenyon, 1999). Mutations in the *C. elegans* Notch receptor *glp-1* that phenocopy germline ablation by blocking germline proliferation, also extend lifespan (Berman and Kenyon, 2006). This is not due merely to the absence of reproduction (Partridge et al., 2005), since ablation of the precursor cells for the entire gonad also cause sterility, but without any effect on lifespan (Hsin and Kenyon, 1999). Although germline signalling does not extend lifespan in all model organisms, a negative correlation has been observed between reproduction and longevity in multiple species (Aguilaniu, 2015).

The lifespan-extending effect of germline ablation in *C. elegans* led to the hypothesis that there may be a specific *germline signal* that is ordinarily countered by an opposing signal from the somatic gonad. However, the exact details of this signal proved difficult to elucidate. In their seminal paper Hsin and Kenyon (1999) identified *daf-16*, part of

the insulin signalling pathway, as one downstream component of the germline signalling pathway, as it was found to be essential for the longevity response to germline ablation. It was later discovered that the longevity response was also dependent on hormone signalling to nuclear hormone receptor *daf-12* (Gerisch et al., 2007, 2001) and an intestinal gene *kri-1* (Berman and Kenyon, 2006), further underscoring the dependence of the germline signal on metabolic regulation. Reductions in TOR signalling do not further extend the lifespan of *glp-1* mutants, and the pathway shares downstream effector mechanisms with that of IIS (Panowski and Dillin, 2009). Further research has found a staggering diversity of pro-longevity processes at work in germline-less long-lived worms, including a reduction in mTOR signalling accompanying an increase in *pha-4* and *skn-1*, alterations in mitochondrial function, and a host of other metabolic changes including increased autophagy, steroid signalling, lipolysis, and lipogenesis (Lapierre and Hansen, 2012; Thondamal et al., 2014).

Indeed, there appears to be a broad and intimate connection between altered processing of lipids and germline longevity. Germline-free worms are unusually fatty (O'Rourke et al., 2009), with notable differences in their lipid composition relative to WT (Amrit et al., 2016; Gao et al., 2017; Steinbaugh et al., 2015). There is no simple correlation between overall fat content and lifespan, since worms under dietary restriction contain less fat, while insulin signalling mutants and germline mutants have more fat, yet all these conditions are pro-longevity. Rather, lipid composition, as well as rate of turnover, could be the most relevant factor (Ackerman and Gems, 2012). Multiple studies indicate that the germline-free longevity phenotype is dependent on both lipogenic and lipolytic pathways (Lemieux and Ashrafi, 2016; Yamawaki et al., 2008, 2010). They are dependent for their longevity on lipases (e.g. *lipl-4*) and fatty acid desaturases (e.g. *fat-6* and *fat-7*) (McCormick et al., 2012). These enzymes desaturate stearic acid to oleic acid, and germline mutants have been found to have higher levels of oleic acid, which has been found to be beneficial for lifespan (Gao et al., 2017; Han et al., 2017). Research in our lab has shown that some aspects of these relevant pathways may be distinguishable temporally, in that their transcriptional activation is occurring at different times (Mains, 2018). Recently, Han et al. (2017) demonstrated that *rsk-1*, a conserved TOR pathway substrate, acts from the germline to limit lifespan via an effect on fatty acid desaturation.

Many lifespan-extending treatments also alter lipid metabolism (Hou and Taubert, 2012), underlining the importance of lipid metabolism for ageing pathways. Lipids act as signalling molecules, as stores for energy metabolism, and as structural building blocks, interconnecting many different aspects of cellular metabolism. Moreover, marked changes in lipid composition and distribution have been observed with age (Ackerman and Gems, 2012; Gao et al., 2017). Lipids also serve as substrates to energy metabolism via mitochondrial function, which we explore next.

1.2.3 Mitochondrial function

Mitochondria are the power-houses of the cell, harnessing an electron gradient to generate ATP by oxidative phosphorylation, thereby providing for the energy requirements of diverse cellular processes. During the course of normal ageing in *C. elegans*, mitochondria usually become fragmented and lose overall volume (Regmi et al., 2014; Yasuda et al., 2011, 2006), with a concurrent loss of electron chain oxygen consumption and ATP production (Braeckman et al., 2002; Houthoofd et al., 2005).

Genetic mutations that prolong lifespan, for example disruptions in insulin-like signalling, have been found to also delay age-associated mitochondrial decline (Brys et al., 2010). Mitochondria are not static organelles, but rather, they dynamically change in shape and size in response to altered energy requirements. Increased levels of elongated mitochondria via mitochondrial fusion have been found to be a necessary (although not sufficient) component of several longevity phenotypes, including *daf-2* mutants (Chaudhari and Kipreos, 2017), and DR has been found to increase mitochondrial network remodelling and flexibility (Weir et al., 2017).

Notably, insulin-like signalling mutants have been found to have higher levels of ATP than WT organisms (Chaudhari and Kipreos, 2017). Fused and elongated mitochondria are more efficient at generating energy, and maintaining high levels of energy at older ages is expected to promote cellular maintenance mechanisms such as protein folding and degradation, which both require high levels of ATP (Chaudhari and Kipreos, 2018).

The relationship between mitochondrial function and cellular health is multifaceted and non-linear. Mitochondrial functioning generates as a byproduct multiple forms of reactive oxygen species (ROS), including superoxide, hydrogen peroxide and others. ROS cause oxidative stress and at high doses lead to cellular damage, but at lower doses they actually promote health via a mitochondrial hormesis effect or ‘mitohormesis’ (Ristow and Zarse, 2010; Schulz et al., 2007; Yun and Finkel, 2014). Intriguingly, hormesis effects may affect different cell types differently, for example, specific hormetic effects on mitochondria in sensory neurons have been reported as leading to an overall positive lifespan impact (Maglioni et al., 2014). Mitohormesis is conserved in mammals (Bárcena et al., 2018), and may underlie the notable failure of antioxidants to yield therapeutic benefits in clinical trials (Bjelakovic et al., 2007).

The mitochondrial unfolded protein response (UPR^{mt}) is a central part of mitohormesis. UPR^{mt} acts as a retrograde signalling pathway which helps coordinate nuclear transcriptional responses to the intra- and extra-cellular environment (Yi et al., 2018), but this essential response becomes less efficient with age. For example, the activity of glyoxalase-1, one of the enzymes responsible for detoxifying mitochondrial ROS, is markedly reduced with age despite unchanged levels of transcription, and over-expression prolongs lifes-

pan (Morcos et al., 2008). Suppression of nutrient sensing pathways enhances lifespan partly by increasing the activities of cellular stress response transcription factors such as *skn-1*, *hsf-1* and *daf-16*, resulting in increases in the transcription of chaperones that help reduce the burden of misfolded proteins and other forms of cellular damage resulting from oxidative stress. The oxidative stress response transcription factor *skn-1* is itself an important longevity modulator implicated in insulin signalling, lipid processing and germline longevity alongside multiple other cellular functions (Blackwell et al., 2015). There is a complex relationship between stress resistance and longevity (Zhou et al., 2011), which is apparent even early in ageing (Ben-Zvi et al., 2009).

In order to optimise cellular energy generation as well as damage control and repair processes, cellular biorhythms coordinate mitochondrial and metabolic processes in order to ensure they are optimally timed (Langmesser and Albrecht, 2006). Indeed, there is some evidence that even in *C. elegans*, complex temporal regulatory dynamics control interweaved cycles of energy generation and repair (Simonetta et al., 2008). Circadian (24-hourly) and diurnal (12-hourly) cycles play a role in *C. elegans* energy metabolism in a manner that is conserved with other organisms, despite the fact that *C. elegans* typically live underground in the wild (Migliori et al., 2011; Zhu et al., 2017).

1.2.4 Epigenetic modifiers are metabolic sensors

There is evidence that many effectors of the epigenetic regulation of organismal processes respond to metabolic factors. While *C. elegans* does not have DNA methylation in the same way that mammals do, the enzymes that modify histone tails, and thereby serve an important function in modulating the cellular ‘memory’ of environmental changes, are dynamically sensitive to intracellular levels of common intermediary metabolites (Gut and Verdin, 2013; Katada et al., 2012). For example, histone methylation is sensitive to levels of *S*-adenosyl-methionine (SAM), histone phosphorylation is sensitive to the ATP:ADP ratio and histone acetylation to acetyl-CoA levels (Katada et al., 2012). Moreover, these processes are themselves dynamic and there is a constantly shifting equilibrium between the enzymes that add, e.g., acetylation marks, and those that remove them, which shifts with environmental changes such as the time of day (Nakahata et al., 2009) and levels of nutrient availability (Wellen et al., 2009).

Histone modifications are known to change during ageing (Pu et al., 2018) and can promote longevity (Alvares et al., 2014; Cascella et al., 2014), although in some cases these results have been questioned. For example, overactivation of the NAD⁺-dependent histone deacetylase *sir-2.1* was widely reported to promote longevity (Tissenbaum and Guarente, 2001), but this result was later found to be dependent on an unreported alteration in genetic background (Burnett et al., 2011). Nevertheless, the close relationship between sirtuins and

other metabolic pathways which influence longevity, such as the mitochondrial unfolded protein response and *daf-16*, continues to be elucidated in greater detail (Mouchiroud et al., 2013).

Histone modifications affect cellular processes by altering chromatin organisation and thereby changing the levels of transcription of genes to activate specific programs, or to guard against adverse transcriptional changes (Pu et al., 2015). For example, mitochondrial stress causes epigenetic changes that activate protective transcriptional programs and thereby promote longevity (Tian et al., 2016). One of the specific histone methylation markers (trimethylation of lysine 4 on histone H3) was recently shown to activate transcriptional programs that led to the accumulation of mono-unsaturated fatty acids (MUFAs), and that these MUFAs were themselves active in extending lifespan (Han et al., 2017). This underscores the complexity of the bidirectional relationship between histone modifications and dynamic metabolic levels. In some cases, such effects can even persist transgenerationally (Greer et al., 2016, 2011), although mechanisms explaining such intergenerational memory are still being elucidated. These interrelationships and the ways that they shift in response to environmental perturbations, temporally, and during the course of ageing, are the subject of ongoing research.

1.2.5 Summary: metabolism and ageing in *C. elegans*

The various interrelationships that have been discussed in this section underscore the complexity of the multifactorial relationship between cellular metabolism and longevity (Figure 1.4).

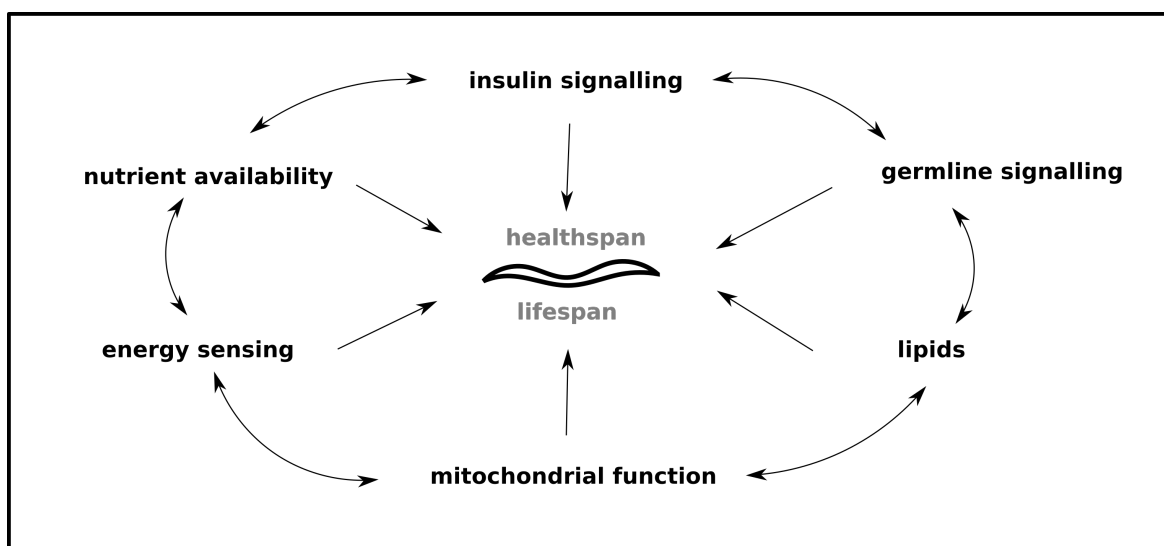


Figure 1.4 Summary of metabolic pathways impacting lifespan and healthspan.

There is a complex balance to be achieved between different cellular and organismal processes, which are responsive to environmental conditions, ultimately determining the temporal sequencing and interplay of different regulatory programs in ways that are either beneficial or detrimental to health over the lifespan. Various metabolic interventions regulate lifespan via a series of partially convergent and overlapping downstream mediators, including *daf-16* (FOXO), *pha-4* (FOXA), *hsf-1* and *skn-1* as well as nuclear hormone receptors such as *daf-12*. A large number of metabolites, lipids, polypeptides and also microRNAs have been implicated as cofactors and regulatory molecules in these pathways – Denzel et al. (2019) offer a recent review.

In the face of these complexities, the field still lacks a clear, unifying theory that explains the impact of metabolism on ageing. There is a need to disentangle complex temporal dynamics and elucidate the specific roles of different metabolites in these temporal changes (Soltow et al., 2010), underscored by findings that not all metabolites that can be food sources or metabolic intermediaries are equally important in contributing to the anti-ageing effects of altered nutrient sensing pathways (Soultoukis and Partridge, 2016). Differences in metabolism of cholesterol and lipids over the lifespan also suggest the importance of independent metabolic pathways and discrepant timings for different metabolites (Morgan et al., 2016).

Yet, detailed time series studies of the metabolic and transcriptional changes that take place during normal and perturbed ageing in *C. elegans* have been rare. One reason for this is that the worms are hermaphrodites and therefore, under ordinary conditions, produce offspring which contaminate ageing samples. Yet, the interventions which are typically used to prevent reproduction, such as mutations in the germline (as discussed above) or treatment with chemicals such as fluorodeoxyuridine (FUdR) have themselves been shown to affect ageing and metabolism (Brunquell et al., 2014; Davies et al., 2012; Feldman et al., 2014). FUdR has been shown to affect metabolism (Davies et al., 2012) and enhance stress response (Anderson et al., 2016). Proteostasis is also enhanced by FUdR treatments independently of germline or gonadogenesis mutations (Brunquell et al., 2014; Feldman et al., 2014). Thus, the study of the temporal and metabolic dynamics of ageing in *C. elegans* needs new approaches.

1.3 Systems biology approaches to the study of metabolism

Recent years have witnessed a paradigm shift from the reductionist study of individual components of biological systems in isolation, towards whole-systems integrative approaches that assess and interpret large datasets in parallel and search for cross-cutting regulatory interactions (Yugi et al., 2016; Yugi and Kuroda, 2017).

Technical advances are making it possible to generate more and more high-throughput data across several technological layers, known collectively as ‘-omics’ (McCormick and Promislow, 2018; Zierer et al., 2015). These include metabolomics, the measure of small molecular cofactors, nutrients and byproducts of cellular metabolism, as well as transcriptomics (a comprehensive measurement of the simultaneous expression levels of genes, as can be determined by counting messenger RNA transcripts), genomics (detection of the presence of common genetic variants across the genome), and others (Valdes et al., 2013). Nevertheless, the capacity for analysis of these rich data resources still lags behind the capacity for data generation.

Systems biology approaches are needed to interpret these remarkably rich data sources and determine their interconnections and regulatory dynamics (Van Assche et al., 2015). Multiple such approaches have been developed, and may be broadly categorised as knowledge-based or unbiased.

Knowledge-based approaches harness pre-existing biological knowledge in the form of, for example, biochemical reaction pathway maps (Kanehisa et al., 2017) and databases of gene function (The Gene Ontology Consortium, 2019). A shortcoming of knowledge-based approaches is that while they can find new connections or interrelationships, they necessarily capture only a fraction of well-understood core biochemistry and thus may be limited in explorations of truly uncharted territory. They are complemented by unbiased approaches, which, in contrast, are able to harness the full measured dataset even of unmapped biomolecular components, and search for patterns based on complex statistical approaches, but may suffer from their predictions being less interpretable. There are several sophisticated methods of unbiased network and causality inference that are applicable depending on the resolution of the dataset, including those based on Bayesian statistics and information theory (Villaverde and Banga, 2014). Recent comparisons of multiple methods for network inference in systems biology have emphasised that while no single method outperforms all others, the application of multiple methods in conjunction with the results compiled in a ‘wisdom of crowds’ fashion gives a better result than any single method applied in isolation (Hill et al., 2016; Marbach et al., 2012).

Machine learning approaches can also be used for this purpose. For example, Barardo et al. (2017b) use machine learning to predict lifespan-extending molecules based on a machine learning predictor built from the DrugAge database (Barardo et al., 2017a), which annotates information about small molecules and their effect on the lifespan of model organisms. Their models used not only the annotated lifespan effects but also associated Gene Ontology (GO) terms from associated enzymes and chemical descriptors calculated from the molecules’ chemical structures. Barardo et al. (2017b) show that the best predictive accuracy is achieved using both biological and chemical features in combination. They were also able to determine that the most significant GO terms include

those related to mitochondrial processes, enzymatic and immunological processes, and metabolic and transport processes. Similarly, Liu et al. (2016) used a network-based approach to predict lifespan-modulating drugs in *C. elegans* based on an annotated set of known lifespan-modulating drugs and a set of drug-protein interactions curated from the literature. Chemical similarity between candidate predictions was used as a validation measure in this prediction. However, as a note of caution, a recent review of machine learning for ageing research highlighted the importance of validation for computational predictions of this sort with wet-lab experiments due to high levels of noise (Fabris et al., 2017).

1.3.1 Metabolomics as a tool to study metabolic changes

Metabolomics can be defined as the comprehensive and quantitative analysis of all metabolites within a biological system (Fiehn, 2002). A wide range of small molecules – which may be intermediary metabolites, common cofactors, xenobiotics or secondary byproducts of metabolic processes – are measured simultaneously from the same sample (Klassen et al., 2017; Liu and Locasale, 2017; Putri et al., 2013). Metabolomics has been called the missing link between genotype and phenotype (Fiehn, 2001).

In contrast to other methods used to measure aspects of metabolism, metabolomics usually refers to the study of a system as a whole, rather than a specific part of that system (Liu and Locasale, 2017). There are a range of different specific types of measurement that are considered metabolomics. Measurements may be very broad, or *untargeted*, generating ‘profiles’ for both known and unknown metabolites, or they may use special standards to detect only a *targeted* subset of known metabolites (Beger et al., 2016; Carneiro et al., 2019). While untargeted measurements can only give relative quantifications, targeted experiments allow for absolute quantification of the concentrations of molecules. Targeted experiments also allow for clear specific molecular identifications. However, targeted experiments cover only a fraction of the metabolome, and are not able to be used for discovery of novel aspects of metabolism. The advent of untargeted metabolomics heralded a new era of discovery in metabolism research, as a staggering number of unknown metabolites were discovered, overturning the perception that most of the metabolic pathways were already known (Zamboni et al., 2015).

There are many technical challenges in metabolomics measurement and it is not yet possible to identify and quantify every single metabolite that may be present in a given sample. The technology used to perform metabolomics measurements is the same as that used in standard chemical identification: mass spectrometry (MS) coupled to liquid or gas chromatography (LC/GC), and nuclear magnetic resonance (NMR). Neither of these technologies is without challenges and limitations. NMR has a low sensitivity but has

less variance between runs; MS is more sensitive, but is not suitable for certain sorts of molecules (e.g. those that are thermally fragile) and suffers from cross-platform variation and batch effects (Liu and Locasale, 2017). Protocols are therefore typically optimised for the particular research question that is being studied.

In summary, metabolomics gives a readout that is closer to the actual micro-physiological state of metabolic processes than any other. However, precisely due to the sensitivity of metabolic changes to variability and environment, it can be more noisy than other -omics layers.

1.3.2 Whole-genome metabolic modelling as an integrative systems biology approach

Metabolism interacts in a complex way with signalling, gene expression and epigenetic factors to control cellular and organismal phenotypic outcomes. Hence there is a need to perform integrative multi-data-source studies in the context of different ageing phenotypes (Hastings et al., 2019b).

Many studies that harness multiple -omics layers use only post-hoc results interpretation to tie the different layers together, without using a truly integrative analysis approach (for a review in *C. elegans*, see Van Assche et al., 2015). For example, Copes et al. (2015) performed both metabolomics and proteomics in ageing *C. elegans*, but interpreted the changes in the proteome and the metabolome largely separately – the underlying biological samples were not linked, and each technological layer was analysed with separate statistics. The resulting changes in the metabolome and the proteome were then combined and contextualised in the accompanying discussion.

Biochemical knowledge, in the form of pathway maps – interconnected series of metabolic reactions – can serve as a backbone for -omics data integration. For example, Gao et al. (2018) interpret the dataset by projecting onto pathways the key alterations associated with ageing from both the transcriptome and the metabolome of *daf-2* and *eat-2* (a *C. elegans* model for DR) mutants.

Whole-genome ('genome scale') metabolic modelling is a knowledge-based approach in which a whole-genome map, or reconstruction, of all the metabolic reactions of an organism are used together with mathematical constraint-based methods to predict metabolic phenotypes under different conditions (Mardinoglu and Nielsen, 2012; Pfau et al., 2011). Prior to 2016, no comprehensive whole-genome model of metabolism existed for *C. elegans*, and then two were published back-to-back (Gebauer et al., 2016; Yilmaz and Walhout, 2016). Subsequently, through a workshop series that was initiated as a part of this project (Hastings et al., 2017), a consensus model (named 'WormJam') has been created and extended (Witting et al., 2018).

This type of metabolic modelling approach will be used in this thesis, together with measured data including time-resolved measures of gene expression and metabolomics. Integrative approaches such as these have the potential to provide insights at an unparalleled level of detail into the specific pathways and reactions affected by, or causing, the observed temporal shifts in metabolism. They allow a whole-systems contextual understanding to be developed of the interrelationships between ageing and metabolism.

Furthermore, various strategies are employed based on this modelling approach to probe the evidence for metabolic regulation through the lens of evaluating the different constraints and flexibilities in the system as a whole (Rosato et al., 2018). This gives information about metabolic regulators in the context of known longevity modulators in order to predict possible interventions.

1.4 About this thesis: Metabolic modelling with multi-omics data in ageing *C. elegans*

A better understanding of the role of metabolic changes and remodelling during ageing necessitates a deeper, temporally sequenced and holistic map of the metabolic landscape during the course of normal ageing. Moreover, the search for metabolic regulators of ageing is of crucial importance in the quest for healthspan-improving interventions, as such metabolic controllers may potentially lead to the discovery of viable supplements (Finkel, 2015).

1.4.1 Objectives

The general objective of my PhD is to disentangle and temporally sequence the key metabolic changes occurring during ageing in *C. elegans*.

To tackle this challenge, I used a time-resolved series of metabolomics measurements in three non-reproductive strains together with metabolic modelling. This study is somewhat different to the previous studies in the strains that were used, and in the focus on more temporally comprehensive measurements during early ageing in order to tease apart the very complex and synchronized metabolic alterations that are happening during early ageing.

In addition, the metabolomics measurements were planned alongside a series of linked gene expression measurements (reported in full in Mains, 2018). This is a rich multi-omics dataset that allows powerful systems biology approaches to interpret changes in the context of the perturbations of the system as a whole that take place during ageing.

1.4.2 Publications

This work has led to the following publications:

1. Workshop report on the first WormJam meeting and subsequent community effort to annotate and extend that model: Hastings et al. (2017).
2. Paper describing community work on developing a shared WormJam model: Witting et al. (2018).
3. Review paper describing multi-omics integration with metabolic modelling in *C. elegans*: Hastings et al. (2019b).
4. Research paper describing metabolic changes in ageing in *C. elegans* and a novel multi-omics integrative modelling approach: Hastings et al. (2019a).

There is some natural overlap in content described in these publications with the relevant sections of this thesis. However, no part of any of the previous publications has been reproduced in full in this thesis. Where appropriate, these publications have been referred to by citations within the text.

All computational methods I have used have been implemented either in R (R Core Team, 2016) or in Python (Rossum, 1995) and are detailed accordingly. Wherever possible, pre-existing libraries and methods have been used; these are named in the appropriate Methods sections.

1.4.3 Organisation of the thesis

The remainder of this thesis is organised as follows.

The next chapter (Chapter 2) describes the analysis of a time-resolved metabolomics dataset to discover the temporal sequence of changes in the metabolome that occurs during normal ageing, including strain-specific and strain-independent features.

After that, in Chapter 3, I describe the WormJam consensus whole-genome model of the metabolism of *C. elegans* that was developed as a community effort to support metabolic modelling in this organism, and my contributions to the development and validation thereof.

The following chapter, Chapter 4, describes an approach to Flux Balance Analysis that harnesses metabolomics information together with transcriptomics, and interprets the resulting series of predicted intracellular fluxes for each time period in the ageing time series.

Chapter 5 describes the results of a series of simulations I performed with the overall multi-omics and modelling pipeline with the objective of determining metabolic regulators of, bottlenecks in, or influences on, the overall ageing process.

The final thesis chapter, Chapter 6, provides a general discussion and contextualisation of the preceding results.

2. Metabolic changes during ageing

2.1 Introduction

In this chapter, I will present the results of our time series metabolomics study which investigated the changes in the metabolome during ageing in *C. elegans*.

Metabolism plays a central role in the ageing process, thus we can expect that the measurable concentrations of metabolites in an organism – the metabolome – will change during the course of ageing. This is indeed what has been observed in *C. elegans* across several previous studies (Copes et al., 2015; Davies et al., 2015; Fuchs et al., 2010; Gao et al., 2017; Pontoizeau et al., 2014; Witting and Schmitt-Kopplin, 2016).

For example, the metabolomic ‘signature’ in conditions of long life has been reported (Fuchs et al., 2010), as well as under dietary restriction (Pontoizeau et al., 2014). The particular changes associated with ageing in normal-lived as compared to long-lived strains have also been reported (Davies et al., 2015). The metabolic and proteomic changes accompanying ageing were assessed together in (Copes et al., 2015). Wan et al. (2017) studied the age-associated changes in the metabolome specific to germline signalling mutants. Table 2.1 gives a summary of these prior studies.

| Study | Genotype(s) | Longevity Pathway(s) | Time points | Method | Other info. |
|--------------------------|----------------------------|--|--|--------------|---|
| Fuchs et al. (2010) | <i>daf-2, ife-2</i> | insulin signalling, translation initiation | 15h, 144 h, 240 h post-hatch | NMR | Offspring removed by mechanical filtering |
| Pontoizeau et al. (2014) | <i>N2, eat-2, slcf-1</i> | Dietary restriction | 1 day, 7 days post-hatch | NMR | FuDR added |
| Copes et al. (2015) | <i>glp-4</i> | Notch receptor / germline signalling | days 4 and 10 | GC-MS | Liquid culture |
| Davies et al. (2015) | <i>N2, daf-2</i> | Insulin signalling | 2.5/3.5 days (N2/ <i>daf-2</i> respectively) and days 6, 8, 10, 13, 16 and 10 post-hatch | NMR | FuDR added |
| Wan et al. (2017) | <i>glp-1, glp-1;daf-16</i> | Notch receptor / germline signalling | days | NMR, UPLC-MS | |
| Gao et al. (2017) | <i>N2, glp-4</i> | Notch receptor / germline signalling | days 1-10 (for N2), 1,3,5,7,9 (for <i>glp-4</i>) | LC-MS | FuDR added (for N2) |

Table 2.1 Ageing metabolomics studies in *C. elegans*

Fuchs et al. (2010) examined the shared aspects of the metabolome of three long-lived *C. elegans* conditions: *daf-2* and *ife-2* long-lived strains, and dauer larvae. They identified 26 metabolites and sampled three time points, and then ranked the features they detected for their consistency across the long-lived conditions. They found particular differences in e.g. trehalose; choline, phosphocholine and glycerophosphocholine (associated with lipid metabolism); acetate, malate and succinate (associated with carbohydrate metabolism), propanoate, NAD⁺ and branched chain amino acids. Some of the concentration differences they observed were *daf-16* dependent, but not all, e.g. trehalose was not.

Davies et al. (2015) compared the metabolome of wild-type worms at seven different time points to that of the long-lived *daf-2* mutants at the same time points, identifying 33 individual metabolites as distinguishing between wild-type and the long-lived condition. Trehalose was found to accumulate in the long-lived condition, while becoming depleted in wild-type. Similar patterns were observed for branched chain amino acids. In contrast, putrescine and tyrosine were found to be accumulating in wild type but not in the long-lived condition. They used their metabolomics time series data to infer that the metabolome of the older *daf-2* organisms was closer to that of 'middle-aged' wild-type worms than wild-type worms of a comparative chronological age.

In a similar comparative study, (Pontoizeau et al., 2014) compared the age-related changes in the metabolome between wild-type and dietary restricted worms, using two genetic models for dietary restriction: *eat-2* (which reduces food intake) and *daf-18* (part of the insulin signalling pathway). They concluded that dietary restriction delays the metabolic changes that occur during normal ageing.

Copes et al. (2015) explored the metabolome and proteome changes between days 4 and 10 of adulthood in liquid cultures of the *C. elegans glp-4* strain. They successfully identified 186 metabolites in their assay, and described those metabolites that were significantly different between the old and young samples, finding significant changes in free fatty acid levels, an increase in sorbitol levels, amino acid changes, and decreases in pyrimidine and purine metabolite levels. Furthermore, they report that supplementation with hypoxanthine and cytidine increased lifespan. The discovery that certain metabolites decrease in level during ageing and are able to extend lifespan when supplemented hints at bottlenecks in metabolism arising due to the ageing process, when cellular processes, for various reasons, are no longer able to replenish metabolites that are needed. Taken together, these results provide strong evidence for metabolic changes not being mere by-products of the ageing process, but closely associated with the causal processes leading to the development of dysfunction.

Wan et al. (2017) explored the age-associated changes in the metabolome in germline free *glp-1* mutants and in the *glp-1;daf-16* mutant which is not long-lived. They found changes in ageing-related metabolites, including increased concentrations of pyrimidine

and purine metabolism intermediates and decreased concentrations of the citric acid cycle intermediates, many of which were not present in the double mutants indicating that they may be related to the germline-free longevity phenotype which is abrogated by concurrent knockout of *daf-16*.

The majority of these ageing-associated studies used just a few time points, e.g. one for old worms and one for young worms, and did not aim to be temporally comprehensive. Recently, Gao et al. (2017) performed a detailed time series of metabolomics measurements, targeting a large number of different metabolites within the categories of fatty acids, amino acids and lipids. They obtained one sample per day for the ten days of adulthood in wild-type worms with FuDR added, and a sample every two days in long-lived sterile *glp-4* worms. However, even this study only obtained at most one measurement per day in their time series.

Our study follows on from these previous studies and extends them in two ways. Firstly, in the bulk of the previous studies, fewer time points have been used compared to what we have assayed in early ageing. For most of the preceding studies (with the exception of Gao et al. (2017)), the timings of ageing-related changes in the metabolome have not been investigated in detail. In contrast, our study sought to investigate the temporal sequence of changes and to differentiate between changes that take place during early ageing and those that take place later in ageing. Secondly, the strains we used allowed us to follow the metabolic changes due to normal ageing while avoiding the contamination by progeny in a way that we believe is metabolically closer to wild-type than the usual protocols used for this purpose, as described next.

2.2 Methods

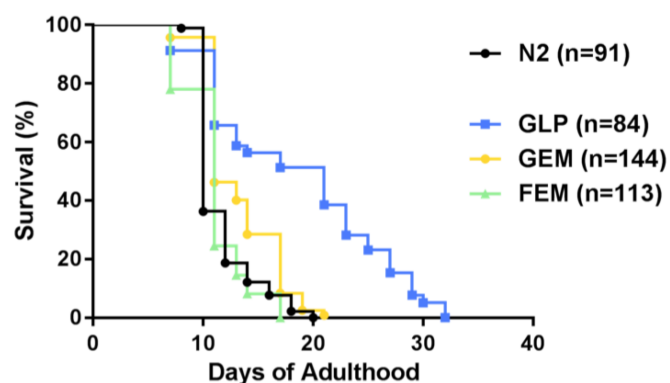
2.2.1 Strains used in the study

The strains used in this study were *glp-1(e2144)ts*, *gon-2(q388)ts;gem-1(bc364)*, and *fem-3(q20)ts* (hereafter referred to as GLP, GEM, and FEM). These are all temperature sensitive mutants that operate only at restrictive temperatures. Conditional sterility was obtained by growing L1 larvae to adulthood at the restrictive temperature (25 °C), which may induce a slight heat stress.

The *glp-1* gene is a popular model for germline longevity: GLP animals possess a complete gonad but no germline, and are long-lived. Many studies of ageing metabolism in *C. elegans* have been conducted in this or other (closely related) *glp* mutants, yet this mutation is known to have pervasive effects on metabolism. Thus, our study also used two other non-reproductive strains, not yet as widely used in ageing research, which are not long-lived (Figure 2.1), and which we believe are metabolically closer to wild type.

Figure 2.1 FEM and GEM strains have lifespans comparable to wild type, while GLP animals are significantly longer lived.

Lifespan analysis of animals at 25 °C. Wild type (N2) have a mean lifespan of 11.51 days. GEM animals have a mean lifespan of 11.71 days. Data shown are representative of three biological replicates. Thanks to Bhupinder Virk for this Figure.



The *gon-2* gene encodes a cation channel required for the development of gonadal tissue (Sun and Lambie, 1997). This mutation is enhanced by the concurrent mutation of the solute carrier *gem-1* (Kemp et al., 2009). Our GEM animals mostly do not reproduce, however, at 25 °C the penetrance of gonadogenesis failure is less than 100%. A highly variable and degenerate gonad that in a very small number of cases results in some reproductive capabilities is visible by day four. The gain-of-function *fem-3* mutation, has 100% penetrance, and causes the germline to become masculine (Barton et al., 1987). Thus, FEM animals possess a normal hermaphrodite gonad but produce no oocytes, only large quantities of sperm and no progeny.

2.2.2 Worm maintenance and sample information

Worms were maintained at 16 °C on nematode growth medium (NGM) with OP50 *E. coli* as food source. Synchronised experimental populations were prepared by washing gravid adults and eggs from plates and bleaching in a freshly prepared solution of 1% sodium hypochlorite and 1 M potassium hydroxide. Eggs were allowed to hatch overnight at 25 °C and all animals were arrested at the L1 stage.

Experimental populations were then placed at 25 °C on HT115 *E. coli* on standard NGM plates containing 50 µg/ml Carbenicillin, 1 mM IPTG, and 10 µg/ml Nystatin until harvesting. For GEM, samples were inspected before harvesting and visibly fertile worms were discarded.

We have obtained two samples per day for the first three days of ageing (post-moulting). This higher density of sampling (two per day) on the earlier days enabled a more detailed investigation of the changes during the very early phase of ageing. Thereafter, we obtained one sample per day on days four, five and ten of ageing. The decision to omit the days between five and ten was based on a pilot study of the temporal changes in the transcriptome, which found low variability in gene expression during these days (Mains,

2018). After day ten, the worms start dying, thus, we did not further obtain samples at later time points.

Samples were harvested at the following hours post-feeding of arrested L1:

- Day 1: Hours 41 and 49 post feeding
- Day 2: Hours 65 and 73 post feeding
- Day 3: Hours 89 and 97 post feeding
- Day 4: 117 hours post feeding
- Day 5: 137 hours post feeding
- Day 10: 257 hours post feeding

The GLP time series of metabolomics measurements is unfortunately incomplete in that only the earlier time points (days 1-3) are available for that strain. Therefore, some of the analyses that are reported below on the topic of overall ageing use only the FEM and GEM strains, while for the detailed study of early ageing it was possible also to use the GLP data.

2.2.3 Metabolomics assay

Targeted metabolomics using Mass Spectrometry was performed by the Northwestern Metabolomics Research Center¹ from a batch containing approximately 2000 worms per replicate. Worms were washed several times and snap frozen in water. Samples were homogenized with a bullet blender at 4 °C, protein precipitated with methanol, sonication, and centrifugation. The supernatant was removed and dried in a vacufuge (Speedvac) at 30 °C. All samples were processed in parallel to avoid batch effects.

Targeted LC-MS/MS metabolomics targeting a list of 210 metabolites was performed on a system consisting of Shimadzu Nexera XR LC-20AD pumps coupled to a Sciex 6500+ triple quadrupole spectrometer operating in MRM detection mode through the Sciex Analyst 1.6.3 software. The system includes a dual column setup with dedicated columns for positive ionization mode and negative ionization mode. The results for each sample are therefore the result of two injections. Metabolite concentrations were quantified using Multiquant 3.0 software in a relative manner. The samples were separated on a Waters Xbridge BEH amide column (2.5 μ m, 130 angstrom, 2.1 x 150 mm) operated in a HILIC regime at 40 °C. Solvent A consisted of 95% water, 3% acetonitrile, 2% methanol, 0.2% Acetic Acid (v/v/v/v) 10 mM ammonium acetate, pH approximately 4.2. Solvent B consisted of 93% acetonitrile, 5% water, 2% methanol, 0.2% acetic acid and 10mM ammonium acetate. Organic solvents and acetic acid were Optima grade from Fisher Scientific USA, ammonium acetate was from Sigma Aldrich. 18.2 MOhm water was from a Synergy UV system by Millipore. Gradient at 0.300 mL/min was as follows: 0 – 3 minutes 95% B, 3 – 8 minutes 95 – 50% B, 8 – 12 minutes 50% B, 12-13 minutes 50 – 95% B 13 – 18.1 minutes 95% B.

¹<https://depts.washington.edu/mmcslu/resources/current-research/metabolomics/>

During the injection on columns of opposite polarity solvent continued at 95% B giving each column approximately 23 minutes of equilibration time. Samples were normalised to total protein content quantified by Bradford assay. The names and details of all measured metabolites are listed in Supplementary Table A.1.

2.2.4 Data pretreatment and quality control

Metabolomics data generated from LC-MS platforms, such as the one used to generate our dataset, often contain missing values (Armitage et al., 2015; Di Guida et al., 2016; Gromski et al., 2014). Almost half of the targeted metabolites from the targeted assay used in our case were missing across all samples in our dataset, perhaps reflecting specific challenges with *C. elegans* sample preparation for metabolomics compared to the other sources of sample material, e.g. the hard cuticle. These rows were discarded in full prior to further analyses, leaving 125 remaining metabolites.

In keeping with the standard practice for metabolomics workflows, I then also removed metabolites that had more than 10% such missing values across all samples (Figure 2.2A). After removing these problem rows (metabolites), there was still one sample that had more than 10% missing values across all metabolites in that sample (Figure 2.2B). However, this sample was from a replicate group (day 10 FEM) that only had two replicates, thus it was *not* removed. 10% is perhaps quite a conservative threshold – many protocols use 20% as the cutoff. However, the dataset was fairly noisy so it was important to use such approaches to reduce noise as far as possible.

The remainder of the missing values were interpolated by replacing each missing value with the mean for that metabolite across all samples, so as not to affect downstream analyses. This is also a standard practice in metabolomics workflows.

The dataset was then transformed using the inverse hyperbolic sine, which is linear for small x while asymptotically approaching $\log(2x)$. In particular, zero values are mapped to zero and all other values are mapped to positive values (which is not true for standard log-transformation).

The data were further normalised by mean-centering and scaling so that every metabolite had a mean of 0 and a standard deviation of 1, rendering the values comparable, again a standard practice in metabolomics data analyses (Di Guida et al., 2016; Xia, 2017).

The dataset was still noisy, so as a further quality control measure I calculated the sample distances between replicates (using ‘1-correlation’ as the distance metric) to obtain an indication of individual sample quality in terms of ‘closeness’ within replicate groups. In cases where there were three available replicates and one of them was significantly removed from the other two, I took this as an indication of lower quality and removed those outlying samples, as indicated by the samples coloured in red and labelled in Figure 2.3A.

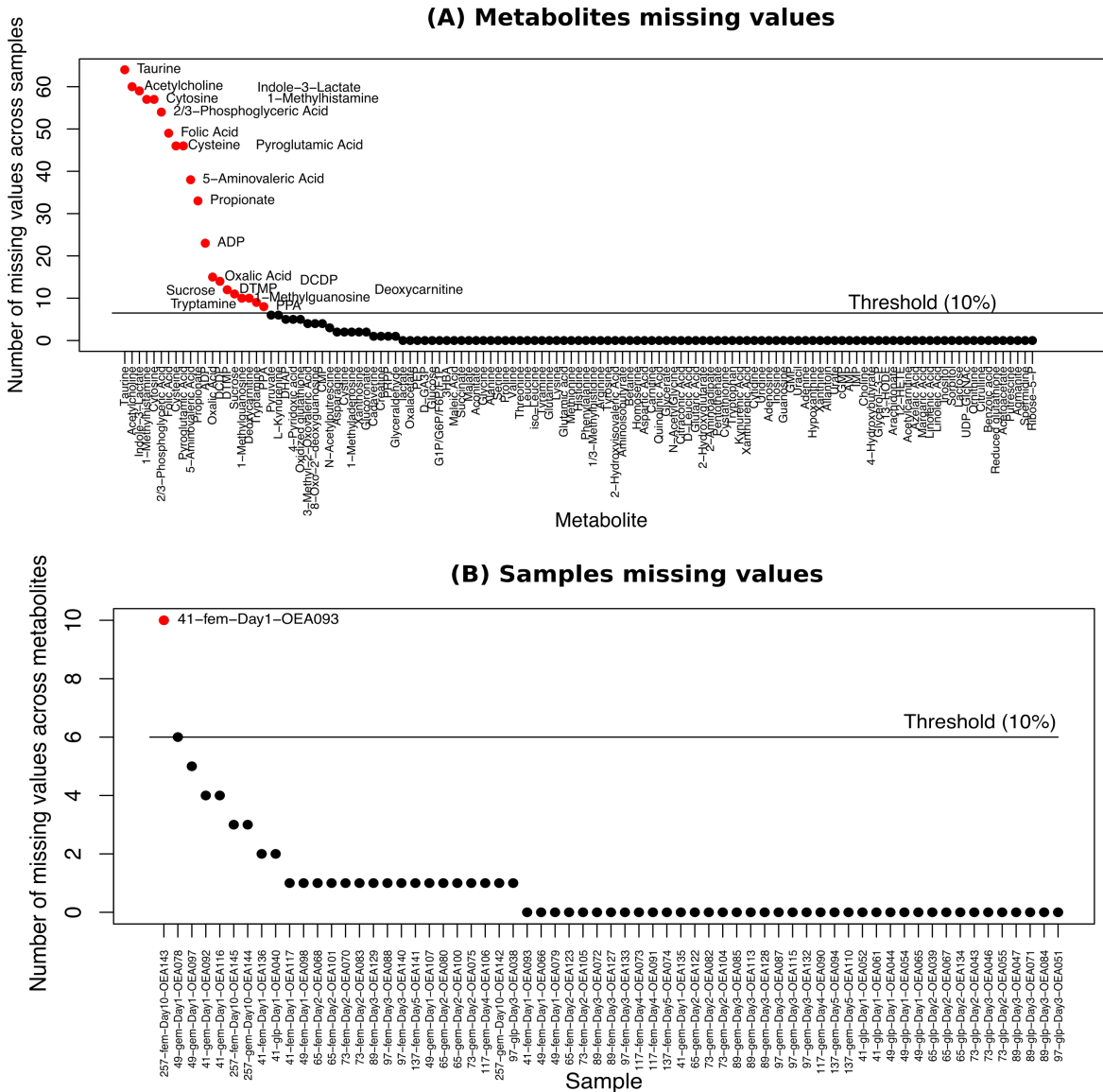


Figure 2.2 Counts of missing values per metabolite and sample.

The numbers shown on the y axis are the counts of (A) samples for which a given metabolite is missing a value across all samples (including GLPs), and (B) metabolites missing values per sample after removing the most problematic metabolites as identified in (A). Each dot indicates (A) a metabolite, and (B) a sample. The horizontal rules in both plots indicate the position of 10% of the values missing, used as a threshold. Those coloured red and labelled have above the threshold, i.e. above 10% missing values.

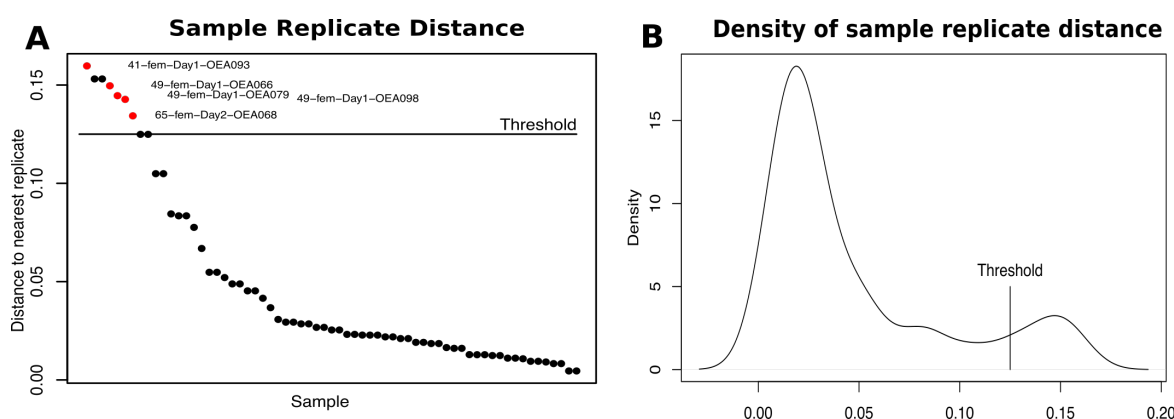


Figure 2.3 Distances to nearest replicate within a sample group.

For each sample in the full dataset (all strains), distance was calculated to the nearest replicate using 1-correlation. Plot (A) shows the distance to the nearest replicate for each sample. Each dot represents a single sample. The samples highlighted in red had an above-threshold distance from the others within the same replicate group, and were thus removed. Only groups which had three replicates available were considered for this quality control step: the dots that are above the threshold but not highlighted in red are in a replicate pair rather than a triplicate. Plot (B) shows the density of nearest replicate distances for all samples, illustrating the choice of threshold.

The choice of threshold for the removal of these samples was based on the observation that the density of nearest replicate distances was distributed such that there was a ‘good’ peak close to zero, and an outlying ‘bad’ peak above 0.10. (Figure 2.3B).

Additional sanity checks that were conducted included computing the sample to sample correlations and distances, and visualising those by main phenotypic sources of variance. Sample to sample correlations were calculated and plotted using the ‘corrplot’ package in R. The density plot of sample to sample distances was plotted by categorizing the pairs (x,y) of samples according to whether they were replicates or had the same age (but different strains) or had the same strain (but different ages). The results of these sanity checks are presented in Section 2.3.1 below.

2.2.5 Multivariate analysis methods: PCA and PLS-DA

Metabolomics datasets are inherently complex, with many possible hidden dependencies between the measured concentration levels of the different metabolites in the assay. Therefore, they are best suited to multivariate rather than single-variable analysis methods (Bartel et al., 2013; Krumsiek et al., 2016), as multivariate approaches are able to use not only the individual contributions of single metabolites but also their interdependencies to explain phenotypic outcomes.

Principal Component Analysis (PCA) is an unsupervised multivariate analysis method that computes a sequence of orthogonal ‘components’ that explain the variance in the dataset, ordered such that the components which make the biggest contribution to explain-

ing the variance appear first while those with the smallest contribution appear last. The PCA returns two matrices for each component: one for scores, containing the positions of samples in the rotated coordinate system, and one for loadings, which gives the metabolite weights. Typically, the first two or three components are able to explain the bulk of the variability in the dataset when taken together, thus PCA is often used for dimensionality reduction. It is a technique that serves as a good starting point for metabolomics data analysis, especially in the exploratory phase prior to the development of specific hypotheses for testing.

To compute the PCA for my dataset I used the ‘prcomp’ package in R. Samples were then visualized as projected into a plane with their positions in the first two principal components, accounting for the majority of variability.

The other multivariate analysis method I have used with the metabolomics dataset is Partial Least Squares – Discriminant Analysis (PLS-DA) (Brereton and Lloyd, 2014, 2018). PLS-DA can be used to infer the metabolites that have the greatest predictive value, which means that they are changing the most in line with changes in the output variable. Technically, PLS-DA is a latent variable regression method based on covariance between the predictors and the response, and it has been shown to efficiently handle datasets with multi-collinear predictors, as is the case of mass spectrometry measurements (Wold et al., 2001). PLS-DA is a linear regression-based method, thus non-linear effects will be overlooked. While PCA is an unsupervised approach, PLS-DA is supervised in the sense that it uses a phenotypic outcome variable (age in hours, in this case) and then attempts to find an optimal arrangement of predictor variables (metabolite values) to predict the given phenotypic values.

To determine which metabolites were significantly changing with age in our metabolomics dataset, I used the PLS-DA implementation in the R package ‘ropls’ (Thévenot et al., 2015), with the sample time in hours as the output (response) variable. As described in Thévenot et al. (2015), this R package implements the PLS-DA approaches with the original version of the algorithm (Wold et al., 2001), including quality metrics which estimate the significance of the model by permutation testing, the permutation diagnostics, computation of the variable importance in projection (VIP) values, and score distances to detect outliers. The VIP values reflect both the loading weights for each component and the variability of the response explained by this component. Thus, they can be used to select those features (metabolites) that are predictive of the response variable (in our case, age in hours) (Pinto et al., 2012). Metabolites in the dataset were considered to be *age-associated* if their VIP given by the PLS-DA model was greater than 1 (which is a standard threshold for this type of analysis).

PLS-DA may lead to overfitting (Gromski et al., 2015), thus, its predictions must be validated. Validation of the PLS-DA model is done by randomising the input values and

comparing the actual predictors to the randomised predictors. The model provides a summary statistic (Q2) for the cumulative predictive performance, and a measure of noise in the underlying dataset (RMSEE) that is the square root of the mean error between the actual and the predicted responses.

2.2.6 Literature and database search for longevity-modulating metabolites

To determine which metabolites were known to be longevity modulators in *C. elegans*, i.e. those that extended lifespan when supplemented to *C. elegans*, I primarily consulted the DrugAge database of known longevity modulators (Barardo et al., 2017a). The DrugAge database also includes metabolites that are detrimental to longevity, and in some cases that led to there being conflicting results in DrugAge (e.g. glucose, for which DrugAge contained one record of an assay in which glucose supplementation had *extended* longevity amid several records of assays in which glucose supplementation had *reduced* longevity). In cases where conflicting reports were available in the database, we went with the majority of reports - for example, in the case of glucose, we accepted the consensus of no positive effect.

A second database resource that was also consulted was the JenAge Ageing Factor Database (Hühne et al., 2014). Furthermore, the published literature was searched for additional longevity modulators using the keywords '*C. elegans*' and 'longevity'/'ageing'.

2.2.7 Pathway enrichment – Fisher's exact test

To determine the relationship between age-associated and longevity modulating metabolites, and to evaluate which pathways were changing the most with ageing, I used Fisher's exact test. This is a statistical method used to determine the likelihood of observing a set of co-occurrences between two groups in a sample of interest. It evaluates this likelihood based on a contingency table of values for two overlapping categories. Fisher showed that the probability of obtaining any such set of values was given by the hypergeometric distribution, as long as the background assignment of values to categories (marginal values) are fixed and there is no association between the categorical values.

I calculated Fisher's exact test using the R method 'fisher.test'.

2.2.8 Determining metabolites with oscillations within days in early ageing

I used another statistical approach, analysis of variance (ANOVA; Harris et al., 2012), to determine which metabolites showed a significant variation within each day during early ageing after controlling for variation attributable to day-on-day changes.

Each time point in the early ageing dataset was assigned to a group which was one of ‘early’ (morning) or ‘late’. These attributes were then passed in to an ANOVA model (‘aov’ in R) together with the other sample attributes, namely hour of collection and strain.

2.3 Results and Discussion

After data pretreatment as described above, the resulting dataset included 105 metabolites and 59 samples, distributed by strain and time point as illustrated in Table 2.2.

| Day | Hours | Sample count (FEM) | Sample count (GEM) | Sample count (GLP) |
|--------|-------|--------------------|--------------------|--------------------|
| Day 1 | 41 | 3 | 3 | 3 |
| | 49 | 2 | 2 | 3 |
| Day 2 | 65 | 2 | 2 | 3 |
| | 73 | 3 | 2 | 3 |
| Day 3 | 89 | 3 | 3 | 3 |
| | 97 | 3 | 3 | 2 |
| Day 4 | 117 | 2 | 2 | 0 |
| Day 5 | 137 | 2 | 2 | 0 |
| Day 10 | 257 | 2 | 2 | 0 |

Table 2.2 Number of replicates per strain and time point for metabolomics dataset.

These time points allowed us to explore in detail the metabolic changes during early ageing, due to the increased temporal density of samples on the first three days, as well as allowing comparisons of young and old samples that allow us to confirm how well our data matches the published literature.

2.3.1 Age is the primary driver of variability in the metabolome

Variability in a dataset may align with experimental factors such as, in our case, age and strain, but may also be due to noise and technical variance. To explore the drivers of variability in the metabolome, I calculated distances and correlations between samples.

The sample to sample correlations, shown in Figure 2.4A, indicate that samples closer in age are more correlated than samples more distant in age. Figure 2.4A also shows that all the samples are quite highly correlated (as would be expected given the tightly controlled sample preparation methods). However, the effect of age was greater than the effect of strain in this dataset: Figure 2.4B shows that the distance between samples (calculated by 1-correlation) are smaller on average for samples with the same age but different strains, than between samples with the same strain but different ages.

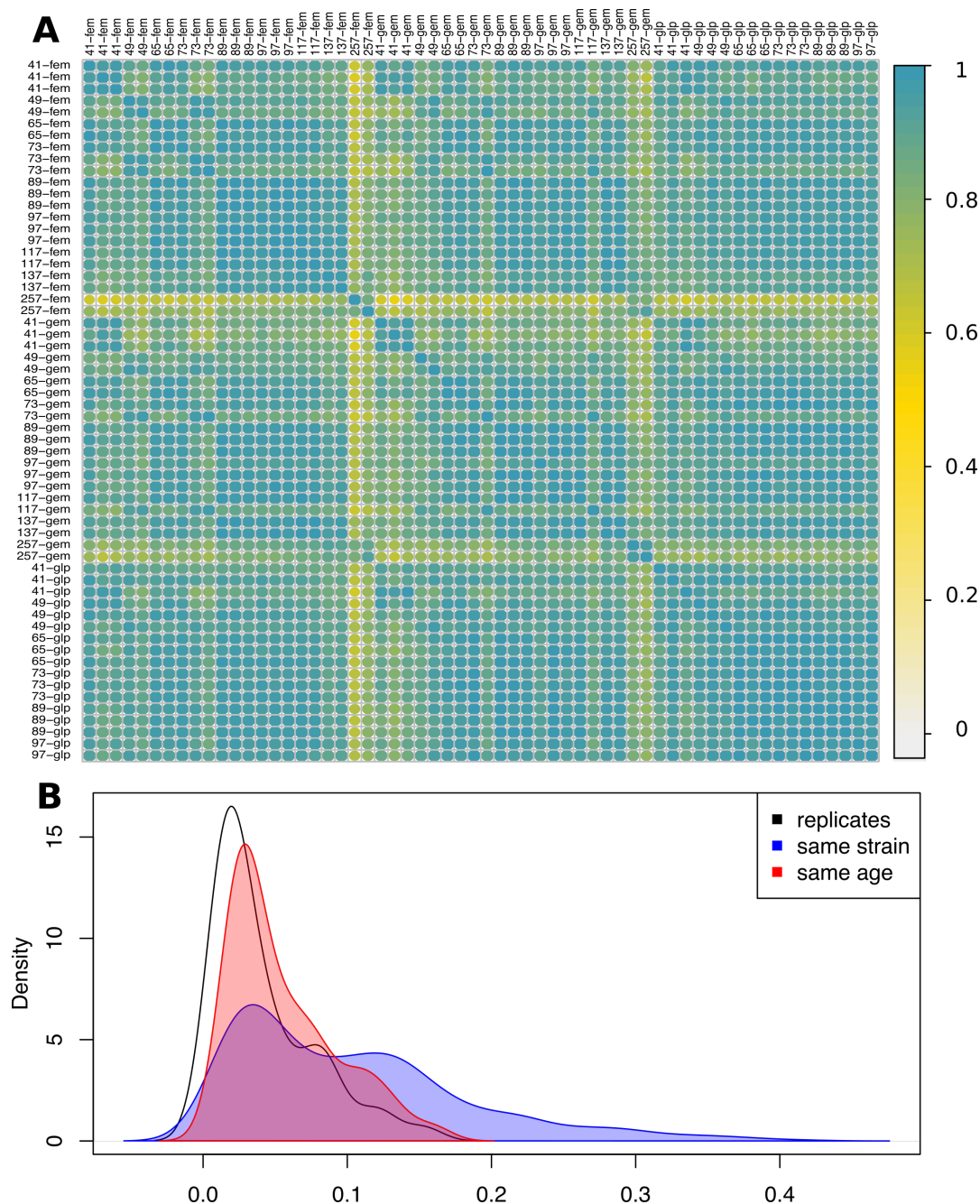


Figure 2.4 Correlation analysis of the metabolomics dataset

The Figure shows (A) the sample to sample correlations as a coloured correlation matrix, and (B) the density plot of sample to sample distances, calculated as 1-correlation, for all samples in our study having the same age (but different strains), and having the same strain (but different ages). In (A), higher between-sample correlations are indicated by a blue colour while lower between-sample correlations are indicated by a yellow colour. In (B), colour indicates comparison group (red=same age, blue=same strain).

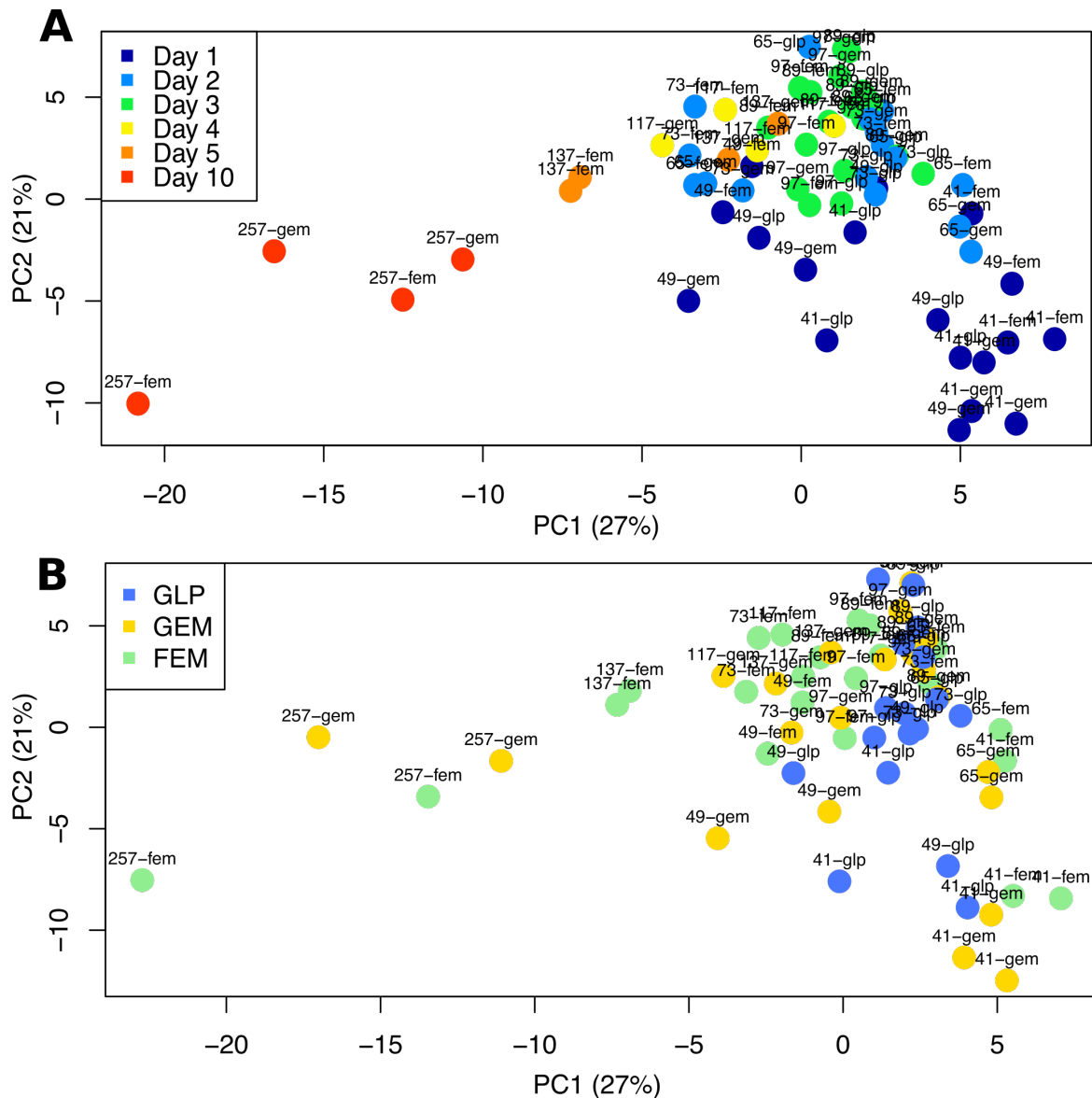


Figure 2.5 Principal component analysis of the metabolomics dataset

The Figure shows the first two principal components accounting for the majority of variability in the principal components analysis of the metabolomics dataset, coloured by (A) day of collection, and (B) strain of sample.

Unsupervised principal component analysis (PCA) allowed further exploration of the main determinants of variability in the dataset. In the visualisation of the first two principal components with samples, coloured by age (Figure 2.5A), it is apparent that the ageing process is the major contributor to variability between the samples, since the first principal component, accounting for the majority of variability, separates young and old samples. In both strains, early ageing and later ageing distribute the samples more broadly (there is a larger distance between samples from the same time point), while middle age seems

more metabolically homogeneous (samples from the same time point are closer together). However, Figure 2.5B shows the same PCs coloured by strain, and in this plot we can observe that strain does not cleanly separate the data along either of the first two PCs. Neither does it do so in the third or fourth PC – illustrated in Supplementary Figure A.1.

This observation was initially surprising, since in our lab (Mains, 2018) we have found that the sample-linked transcriptomics dataset differs more by strain than by age for the same strains and time points, although age does align with the first principal component in this dataset as well. One possible reason for this observed discrepancy in drivers of sample variability between -omics layers is that the targeted metabolomics assay we used covers a small subset of metabolites of central importance, which are likely to be highly conserved across strains under similar environmental conditions. Conversely, transcriptomics is a whole-genome approach that samples thousands of transcripts at a time, and is therefore likely to be much more sensitive to strain-to-strain differences. We might thus hypothesize that the metabolomics data would show more between-strain differences if we had used an untargeted metabolomics assay in which the dataset comprised all metabolites.

An alternative possibility is that the largest source of metabolites in worms is the intestine, whereas the germline may have fewer metabolites but be a large source of transcripts. Thus, the different germline composition of GLPs, GEMs and FEMs may explain the transcriptomics strain difference. However, the strain difference was still apparent even when genes known to be enriched in the germline were excluded (Mains, 2018), making this less likely.

Finally, a plausible reason for the discrepancy is that there are several layers of regulation between transcription and the end result in cellular state as reflected in the concentrations of metabolites. The misalignment between the metabolomics and transcriptomics sources of variability may indicate an important layer of post-transcriptional control of metabolism. In support of this interpretation, we observe that it has been known for more than a decade that gene expression levels for most of the genome do not change much during the course of ageing in *C. elegans* (Johnson, 2013); e.g., Lund et al. (2002) reported that only 1% of the genome shows significant changes of expression during ageing. By implication, therefore, the physiological changes that represent the ageing phenotype in *C. elegans* must be to some extent independent of transcriptional changes. In line with this, Morcos et al. (2008) specifically show that the activity of glyoxylase-1, a metabolically important enzyme, is ‘markedly’ reduced with age despite the fact that mRNA levels remain unchanged. This finding underscores the importance of studying the metabolome itself, as it is closer to the ageing phenotype.

We next explored metabolite concentration changes over the full course of ageing. This next part of our analysis uses only the GEM and FEM strains, as the full time course was

not available for the GLP strain. The GLP abbreviated time series, and accompanying between-strain comparisons, will be reported later in the chapter.

2.3.2 43 metabolites are significantly changing with age

We used PLS-DA to reveal those metabolites that are significantly associated with the age in hours of the sample across all time points (GEM and FEM strains only). As PLS-DA is a supervised method and tries to find the best projection of variability in the dataset to match the response variable, it is important that the result is validated by permutation testing. Supplementary Figure A.2 shows the resulting model scores, loadings and validation statistics. Using a threshold of 1 as a cutoff for VIP as described in the Methods, this approach distinguished 43 metabolites with a significant change in metabolite concentration over the course of ageing, which we will therefore consider 'age-associated'. There are two provisos to this analysis. Firstly, it should be noted that this result applies across the two strains and thus does not identify metabolites that may be changing with age in only one strain – strain-specific differences will be presented in a subsequent section below. Secondly, it should also be noted that PLS-DA is a linear approach, thus non-linear changes with respect to ageing will be missed – although some of these will be discussed in the final section of this chapter.

The fold changes for comparisons between young samples and old samples (days 1 and 10 respectively) are illustrated for these age-associated metabolites in Figure 2.6. The comparison was calculated separately for the two day 1 time points (41 hr and 49 hr) and for the two strains FEM and GEM. Oxaloacetate, an energy-related intermediate of the TCA cycle, is the metabolite that decreases the most in concentration with advancing age, while cadaverine, a diamine, increases the most. These are as might be expected, since the TCA cycle is known to decrease in efficiency with age, while cadaverine is a byproduct of amino acid degradation processes.

Most of these age-associated metabolites display a similar fold change in FEM and in GEM strains. Among the metabolites that show a more distinct strain-specific fold change are guanosine, a purine nucleotide that decreases with age more in FEM than GEM, asparagine, an amino acid which decreases more with age in GEM than FEM, and xanthosine, a nucleoside intermediary in purine metabolism, which increases more with age in FEM than GEM.

The plot reveals mostly subtle differences in fold changes comparing the earlier Day 1 (41 hr) time point to the older samples and the later Day 1 (49 hr) time point to the later samples, indicative of gradual age-associated changes. However, some metabolites show a larger discrepancy between the Day 1 comparison groups. Metabolites showing a larger discrepancy between the comparison groups include guanosine, cytidine, uridine,

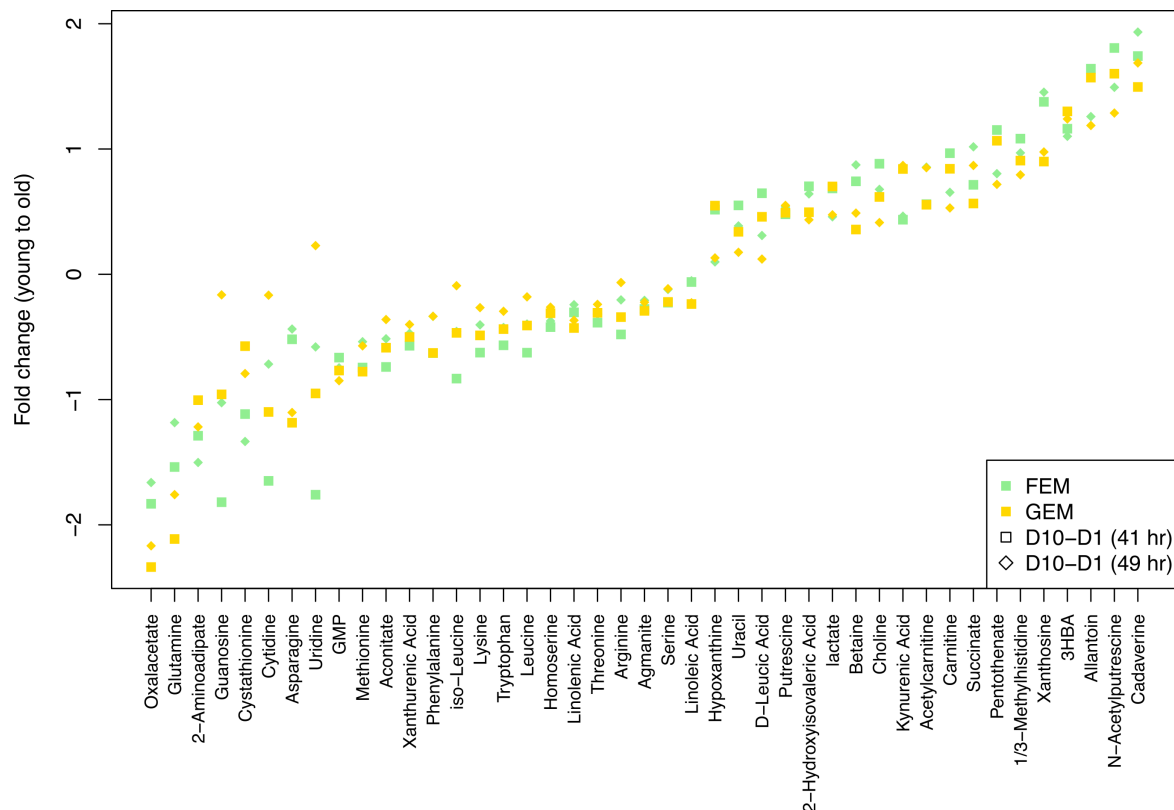


Figure 2.6 Fold changes comparing young to old concentrations for age-associated metabolites. The Figure illustrates the log fold changes from day 1 samples (two hour groups, 41 and 49) to day 10 in our dataset, for GEM and FEM strains, calculated as $\text{mean(D10)} - \text{mean(D1)}$ for each D1 hour group and each strain. Squares indicate the first time point on D1 was used for the comparison (41 hr), while diamonds indicate the second time point on D1 was used for the comparison (49 hr). There was only one time point on D10 (257 hr). Colour indicates strain (green=FEM, yellow=GEM). Metabolites are ordered from smallest to greatest average fold change across all comparison groups.

all nucleotides, and allantoin, a byproduct of purine catabolism. The existence of at least some large discrepancies emphasise that the exact choice of young reference point can be of paramount importance in the comparison of young and old samples for the purpose of understanding ageing-related changes.

On average, the D10-D1 (49 hr) comparison showed a smaller absolute fold change than the D10-D1 (41 hr) comparison, as might be expected if the age-related change is gradual and linear (i.e. in Figure 2.6, diamonds are closer to zero than the squares). Exceptions to this pattern are 2-aminoadipate, cystathionine, betaine, succinate and cadaverine. These variations hint at the possibility of non-linear dynamics with respect to time, explored further below.

Of these age-associated metabolites, several recapitulate what is already known about metabolic changes in ageing *C. elegans*. The relationship to previous findings for each of these age-associated metabolites is given in Supplementary Table A.2. Discordant changes

are highlighted in italics in the Table. For example, among the metabolites in our dataset that decrease the most with age, glutamine, guanosine, cystathionine and cytidine were also found to be decreased with age in previous studies. Intriguingly, oxaloacetate was found to increase rather than decrease with age in the long-lived germline-free *glp-1* strain by Wan et al. (2017), although not in their normal-lived double mutant *glp-1;daf-16*, suggesting that oxaloacetate may be beneficial for longevity. Similarly, in the same study asparagine (another of our metabolites that decreases with age) was also found to be increased in the long-lived mutant but not in the normal-lived mutant. On the other end, those that increase the most with age in our dataset, we find that the levels of succinate were decreased with age in two previous studies (Davies et al., 2015; Wan et al., 2017) and allantoin decreased with age in the wild-type strain, but not the long-lived strain, in Wan et al. (2017). However, most of the metabolites that are increasing with age in our dataset, such as cadaverine, xanthosine, 3-hydroxybutyrate and N-acetylputrescine, have not previously been reported in *C. elegans* ageing metabolomics studies.

This comparison to the literature illustrates one of the challenges with this type of metabolic studies: incompleteness. Unlike whole-genome transcriptomics, targeted metabolomics is inherently selective, with just a small number of metabolites reported per study relative to the size of the whole metabolome, which may be tens of thousands of individual molecules (Witting et al., 2018). No single published study yet provides a comprehensive overview of all the metabolites in *C. elegans*.

Furthermore, there are clear between-strain (even within the same study) and between-study differences, which may be due to underlying strain differences or experimental conditions such as the exact timings at which samples were collected for comparisons or the addition of FuDR. There is a big range in timings of sample collections in previous studies, with e.g. (Davies et al., 2015) focusing on older *C. elegans* by collecting young adult (YA, equivalent to our Day 1), day 6, 8, 10, 13, 16 and 20 (although, their days are counted post-hatching, rather than post-final-moult, again rendering exact comparison between studies more difficult), while (Copes et al., 2015) compare only days 4/5 and 12, and (Wan et al., 2017) uses only YA and day 10. Gao et al. (2017) report a comprehensive time series with four larval stages and days 1 to 10, but then do not provide the details of statistical comparisons between specific early and late time points, which might have eased the comparison to other related studies. Lastly, few of the previous studies report the exact timing in hours to which each of their sample days of collection corresponded.

2.3.3 Polar amino acids change the most with age

We wondered whether specific pathways were more enriched for age-associated metabolites in our dataset than others. Using the pathway annotations associated with the metabolite as-

say (Supplementary Table A.1 in the Appendix), we grouped the metabolites into pathways and calculated an over-representation score (using Fisher's exact test) for each grouping, against the background of all metabolite-pathway annotations in the metabolomics dataset, to determine whether that pathway included more age-associated metabolites than would be expected by chance, given the size of the pathway and the size of the overall dataset. This metric is obviously dependent on the selection of metabolites from each pathway included in the dataset, over which we had no control, and which cannot be assumed to be representative or unbiased (which would be prerequisites of a robust associated statistics). However, this metric can still be used exploratively to compare the pathways within the dataset.

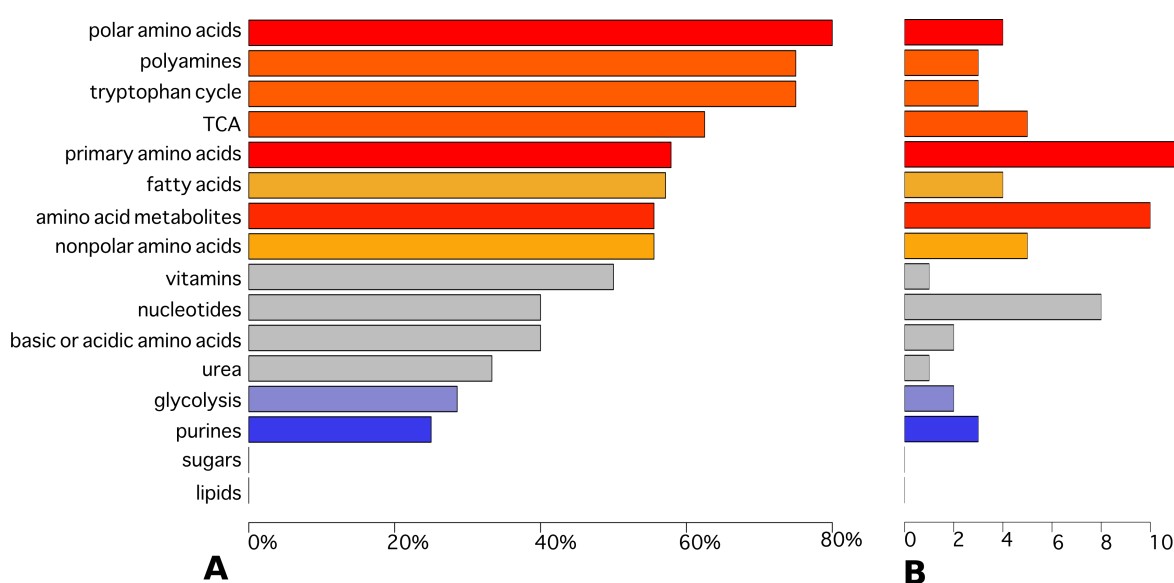


Figure 2.7 Pathway enrichment for age-associated metabolites.

(A) Ranking of pathways by their percentage of age-associated metabolites, coloured by enrichment score. (B) The same pathways and colours, but with the bar sizes indicating the absolute number, rather than percentage, of ageing-associated metabolites in each pathway. Colour shows enrichment for age-associated metabolites (red=over-representation, blue=under-representation).

Figure 2.7A shows the ranking of pathways in our dataset for the percentage of age-associated metabolites in each pathway, while Figure 2.7B shows the number of age-associated metabolites in each of those pathways. The colour of each bar indicates the enrichment score. Metabolites may belong to multiple pathways (listed in full in Supplementary Table A.1).

Polar, primary amino acids are the most enriched, the primary amino acids and downstream metabolites of amino acids are highly enriched and have the largest numbers of age-associated metabolites in the class. Note that a single metabolite may count in several different pathways, for example, all the polar amino acids are also members of the broader primary amino acids pathway. TCA cycle metabolites, polyamines, and metabolites

involved in the tryptophan cycle are also enriched for ageing-associated changes, while the metabolites in the purines and sugars pathways are not changing with ageing in our dataset. Fatty acids show moderate enrichment, and other lipids none, but it should be noted that only a very small number of lipids were detected in our assay, and that in general, specialised methods are typically required in order to measure the concentration of specific fatty acids and lipids in biological samples (Gao et al., 2017; Willenberg et al., 2015; Witting and Schmitt-Kopplin, 2016).

Our overarching interest is not only in the metabolic *changes* during the course of ageing, but also the possibility of metabolic *regulation* of the ageing process. Therefore, we next considered whether in our dataset there was evidence for a relationship between changes in concentration levels over the course of ageing and having a causal role in ageing – i.e. being a longevity modulator.

2.3.4 Age-associated metabolites are enriched for longevity modulators

We wondered whether the metabolites that change the most over the course of ageing might be known to influence the ageing process in any way when supplemented. As described in the Methods, we consulted databases and the literature to determine which of our detected metabolites from the full dataset had previously been found to extend longevity when supplemented to *C. elegans*. The results of this investigation for our dataset of metabolites are included in Supplementary Table A.3 in the Appendix, while a summary of the overall counts is shown in Table 2.3.

| | Age-associated | Not age-associated |
|--------------------------|----------------|--------------------|
| Longevity modulating | 25 | 15 |
| Not longevity modulating | 18 | 47 |

Table 2.3 Numbers of metabolites in age-associated and longevity modulating groups.

In total, 40 of our total 105 measured metabolites were associated with beneficial effects on lifespan in the examined databases and literature, and 25 of those – more than half – were amongst the 43 age-associated metabolites. This is more than would be expected if the 40 were distributed according to chance: Fisher’s exact test for category over-representation for the contingency table illustrated in Table 2.3 gives $p < 0.0005$. This relationship is not unexpected, but has not been explicitly examined previously in a data-driven fashion.

We explored this relationship more deeply by interpreting it in the context of pathways. For each pathway, we calculated the percentage of age-associated metabolites and the percentage of longevity modulators. The results are shown in Figure 2.8. In this bubble plot, pathways in which ageing-associated metabolites are over-represented are shown in red, while those in which ageing-associated metabolites are under-represented are shown

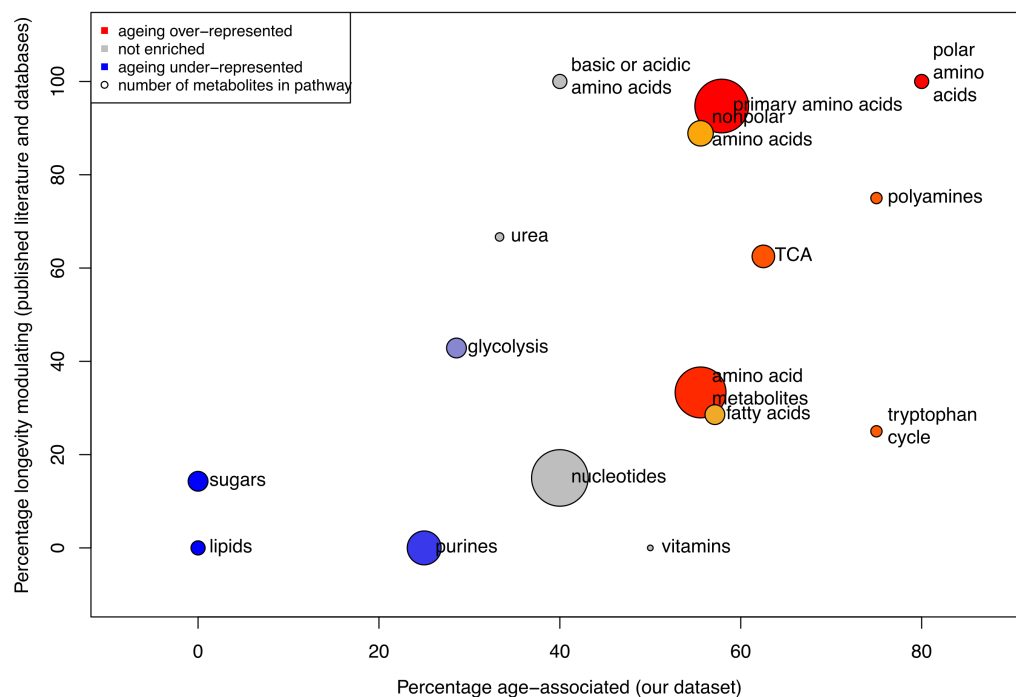


Figure 2.8 Bubble plot for pathway enrichment.

Each bubble represents a pathway. The bubble size shows the number of metabolites in that pathway included in our dataset. Colour shows enrichment for age-associated metabolites (red=overrepresentation, blue=underrepresentation). Position along the x axis shows the percentage of metabolites in the pathway that were determined to be age-associated, while along the y axis is the percentage annotated as being longevity modulators.

in blue, as per the pathway enrichment score illustrated in Figure 2.7. Circle (bubble) size shows the number of measured metabolites in the pathway. The axes show the two different measures: the percentage of ageing-associated metabolites in the pathway from our study on the x axis, and the percentage of metabolites in the pathway associated with longevity in databases and literature on the y axis.

Figure 2.8 indicates that these measures are broadly correlated on the pathway level. However, discrepancies are also apparent. It is evident that the amino acids have been well-studied as a group for their influence on ageing (e.g. in Edwards et al. (2015)), since they are represented highly on the y axis of Figure 2.8. However, the TCA cycle, polyamines, tryptophan cycle and downstream metabolites of amino acids are lower in the y axis (percentage of known longevity modulators) but still high in the x axis (age-associated in our dataset). It is possible that these pathways represent opportunities for the discovery of novel longevity modulators.

2.3.5 Temporal patterns over the full course of ageing

The dynamic pattern of temporal changes is different for different groups of age-associated metabolites. To explore the non-linearity of the changes further, we used a clustering-based analysis of the subset of the dataset for the age-associated metabolites. The scaled mean values per time point (taking the two strains, FEM and GEM, together) for these metabolites are illustrated in the heat map shown in Figure 2.9A. This heat map has been clustered algorithmically, from which eight main clusters have been selected by visual inspection (Figure 2.9B-I).

The first five of the clusters include metabolites that decrease with age, while the remaining three clusters include metabolites that increase with age.

The first cluster (Figure 2.9B) includes three metabolites that decrease with age between days 5 and 10, which are also the three metabolites showing the most negative fold change between days 1 and 10 (Figure 2.6). This cluster contains the TCA cycle metabolite oxaloacetate and the amino acid glutamine, involved in energy maintenance and protein synthesis respectively, both of which are known longevity modulators. 2-aminoadipate is an intermediate of a lysine synthesis pathway, and although not much is known about this metabolite in *C. elegans*, it has been implicated as a biomarker for human diabetes risk (Wang et al., 2013).

Clusters two and three (Figure 2.9C and D) contain a large group of mainly amino acids that show fluctuations of alternate increases and decreases within a range during early ageing (which will be discussed in greater detail below), then begin to decrease during 'middle' age (days 4 and 5), and continue to decrease gradually between days 5 and 10. The overall pattern of decrease of amino acids with age confirms a well-known imbalance in amino acid metabolism with age in *C. elegans* that has previously been attributed to the change in cellular volume with age (Swire et al., 2009). As *C. elegans* somatic cell count is fixed from the beginning of adulthood (adult animals are post-mitotic), yet the organism continues to grow, cellular volume increases, causing a change in the ratio of hydrophilic to hydrophobic surfaces, which necessitates a decrease in hydrophobic amino acids in order to maintain membrane surface to cell volume ratios (Swire et al., 2009). This phenomenon was also mentioned in Copes et al. (2015); Gao et al. (2017); Wan et al. (2017). However, in our lab we have found that there is a significant change in the worm size from days 1 to 5 (Figure 2.10), but not between days 5 and 10, meaning that in partial contrast to received wisdom, the decrease in amino acid levels later in ageing cannot be due to body size changes. An alternative explanation could be that this decrease relates to the reduction food intake which is one of the known pathologies of ageing in *C. elegans*, as pharyngeal pumping becomes inefficient (Russell et al., 2017) and the intestine atrophies (Ezcurra et al., 2018).

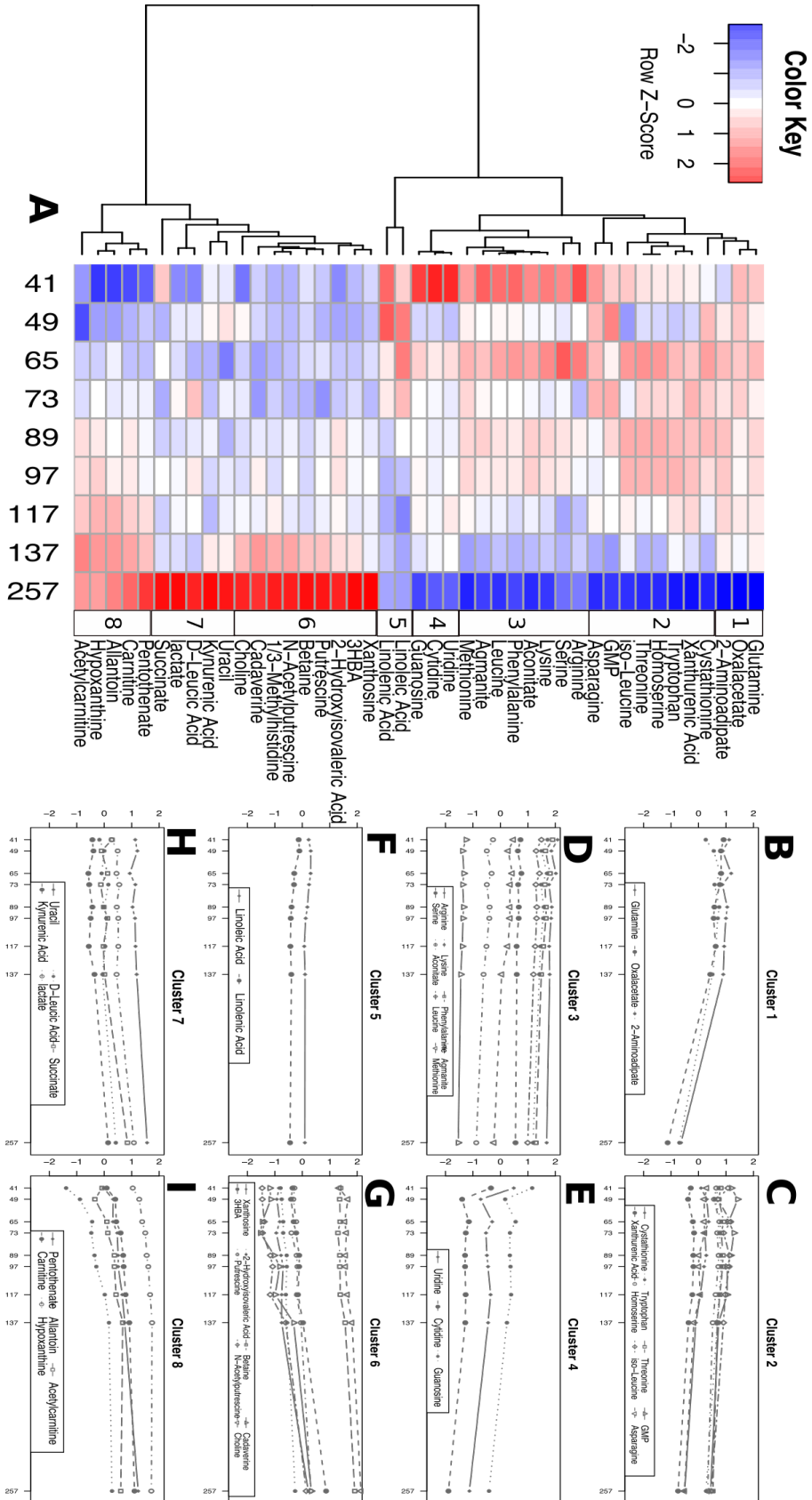


Figure 2.9 Metabolite intensities clustered by hour of collection.

(A) Row-clustered heat map of the mean (across both FEM and GEM strains) intensity values for age-associated metabolites per time point. Cluster numbers are indicated next to the metabolite names. (B-I) The same values shown as separate line plots of the mean intensity for each metabolite per time point, arranged in the eight separate clusters identified in the heat map.

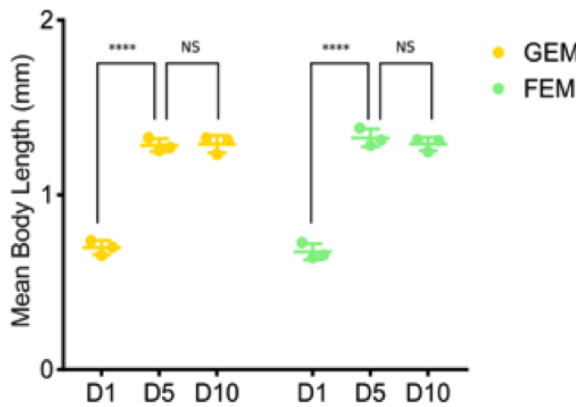


Figure 2.10 Body size (length) at days 1, 5 and 10.

Body length significantly increases between Day 1 (41 hours) and Day 5 (137 hours) in both the GEM ($P < 0.0001$) and the FEM ($P < 0.0001$) strains. Between Day 5 and Day 10 (257 hours) no significant change in body length is observed in either strain. Points plotted represent the mean body length measured from at least 30 animals, biological replicate count $n=3$ for each timepoint, per strain.

This Figure appeared in the Supplementary Material of (Hastings et al., 2019a). Thanks to Bhupinder Virk for this Figure and associated data.

Cluster six provides further support for this hypothesis. Cluster six metabolites (Figure 2.9G) show significant increases between days 4 and 5, and a further steep increase between days 5 and 10. This group contains several polyamines – cadaverine, putrescine and N-acetylputrescine. Polyamines are downstream products of amino acid metabolism and have many cellular functions. Supplementation of putrescine increases lifespan in worms by 10-20% (Edwards et al., 2015), whilst the role of cadaverine remains unexplored. Cadaverine is a precursor of spermidine, which has been associated with lifespan extension by activating autophagy (Eisenberg et al., 2009; Minois, 2014; Minois et al., 2014). Although spermidine was available in our dataset, the levels were too variable between replicates to detect a clear pattern with respect to age.

Autophagy is a well described fasting response that recycles cellular components to restore the energy balance. It is interesting to note that the levels of 3-hydroxybutyric acid (3HBA) behave similarly to the polyamines (i.e. this metabolites is also in cluster 6). 3HBA is metabolised from ketogenic amino acids and/or lipids, and it is a component of ketone bodies, which supply energy during periods of fasting (Veech et al., 2017). At the same time, we observe significantly increased levels of spermidine precursors. Thus, polyamine synthesis may be enhanced in older animals so as to stimulate autophagy when nutrient availability from food intake becomes limiting. Autophagy can also mediate the conversion of intestinal biomass into yolk, causing documented early ageing pathologies (Ezcurra et al., 2018). Therefore, autophagy can have positive and negative consequences for longevity. The seemingly contradictory nature of these two observations might be further clarified by gaining a deeper understanding of the underlying molecular mechanisms that trigger autophagy and the physiological contexts where it is used. A plausible scenario is that the increase in polyamines that may result from reduced food intake at day 10, may promote a switch in autophagy away from lipoprotein pool production into a survival strategy aimed at coping with malnutrition, which may be beneficial for survival. FEM and GEM

level changes are similar for these metabolites, indicating that the age-related changes we observe are independent of reproduction.

Cluster four (Figure 2.9E) contains the nucleotides uridine, cytidine and guanosine, each of which shows a striking decrease during day 1, followed by somewhat of a recovery and then a gradual decrease through to day 10. Cluster five (Figure 2.9F) contains both of the medium-length fatty acids that were included in the set of measured metabolites, each of which shows a small and steady decrease over the course of ageing, in keeping with what was previously observed in Gao et al. (2017).

Cluster seven (Figure 2.9H) contains metabolites that are involved in central carbon metabolism, including the TCA cycle metabolite succinate and anaerobic metabolism byproduct lactate, increases of which are usually associated with excess exercise or other anaerobic conditions such as hypoxia. An imbalance in the TCA cycle in 10-day old animals has been previously described (Wan et al., 2017). These metabolic changes may reflect the widespread fragmentation of mitochondria that begin at around this stage (Regmi et al., 2014; Yasuda et al., 2006). Interestingly, succinate, an important metabolite for multiple convergent longevity-related pathways (Tretter et al., 2016), displays a steep *decrease* in early ageing between the two day 1 time points, then remains relatively stable before increasing between days 5 and 10. This complex trajectory with non-linear behaviour and variability at earlier time points may explain some of the contradictory findings with respect to whether or not succinate increases or decreases with age (Supplementary Table A.2), as if this pattern is typical, the level change attributed to ageing would then depend strongly on the exact timing of the earlier sample used for comparison.

While the other clusters of increasing metabolites show their steepest increase between days 5 and 10, cluster 8 (Figure 2.9I) contains metabolites that increase more steadily across the time series and show significant increases already during early ageing. In this group we have energy-related metabolites carnitine, acetylcarnitine and pantothenate as well as intermediaries of nucleotide metabolism allantoin and hypoxanthine. Interestingly, allantoin has been earmarked as a calorie restriction mimetic (Calvert et al., 2016), indicating that it interacts with and potentially represses insulin signalling, leading to improvements in healthspan and lifespan when supplemented. Copes et al. (2015) found a small but significant lifespan benefit when supplementing hypoxanthine, and it is known to be active both in energy metabolism and in nucleotide synthesis.

2.3.6 Strain differences and patterns during early ageing

Thus far, we have considered the overall pattern of changes during ageing from early adulthood to old age. However, our time series included a higher temporal resolution during early ageing specifically with the intent to tease apart early metabolic changes that

may be related to the onset of age-related pathologies (Ezcurra et al., 2018), and the loss of the normal ability to mount a stress response that occurs at approximately day 2 of adulthood (Ben-Zvi et al., 2009). For these early ageing time points, all three strains were available, which allows us also to compare the metabolism between the GEM and FEM strains, both of which do experience a collapse in stress response, and the long-lived GLP strain in which the collapse is delayed until after day 4 of adulthood.

Firstly, we conducted another PLS-DA analysis on the early ageing dataset, this time using strain as the determining output (response) variable. The model statistical details are shown in Supplementary Figure A.3 in the Appendix. In contrast to the PLS-DA model built with hour of sample collection as response (Supplementary Figure A.2), this model is not robustly predictive (the associated statistic, Q^2 , is close to 0, as illustrated in Supplementary Figure A.3) as there is no clean separation between the strains. However, this analysis nevertheless provides an indication as to which metabolites show between-strain differences in concentrations in at least some of the early ageing time points. Using a VIP cutoff of 1, 35 such metabolites were determined. The full set is illustrated in Figures 2.11 (clustered heat map) and Supplementary Figure A.4 (individual mean concentration level plots). A selection of the concentration level plots showing metabolite scaled intensity values is illustrated in Figure 2.12.

One cluster in the heatmap contains metabolites generally elevated in FEM relative to the other strains, including glycerate, lactate, D-Leucic acid, 3HBA and acetylcarnitine (Figure 2.12A). Glycerate and lactate are indicators of glycolysis, D-Leucic acid is a downstream metabolite of leucine degradation, 3HBA is a ketone body downstream of fatty acid breakdown, and acetylcarnitine is an intermediary in the mitochondrial fatty acid oxidation. Taken together, this cluster of metabolites indicate that the FEM strain may be metabolizing diverse energy sources at a higher rate than the other two strains.

Other metabolites showing a strain-specific difference appear specifically depleted in GEM, including hypoxanthine and cystathionine, particularly during the first time points (day 1-2) in the early ageing time series (Figure 2.12B and C). Hypoxanthine is a purine derivative that is a known longevity modulator, and cystathionine is an amino acid intermediary related to energy metabolism. By contrast, another purine derivative, inosine (Figure 2.12H), is elevated in early ageing in GEM and FEM compared to GLP, although the trend is a decrease with time in all three strains.

A cluster of metabolites are elevated in GLP in particular, including AMP and reduced glutathione (Figure 2.12D and E). Glutathione is the major intracellular redox buffer, and depletion of glutathione in *C. elegans* has been found to have a hormetic effect, extending lifespan at low doses but detrimental at higher doses (Urban et al., 2017). Since the levels of this metabolite start out at the same level for all strains then reduce in GEM and GLP during day 2, it might be hypothesised that this shift is somehow related to the general

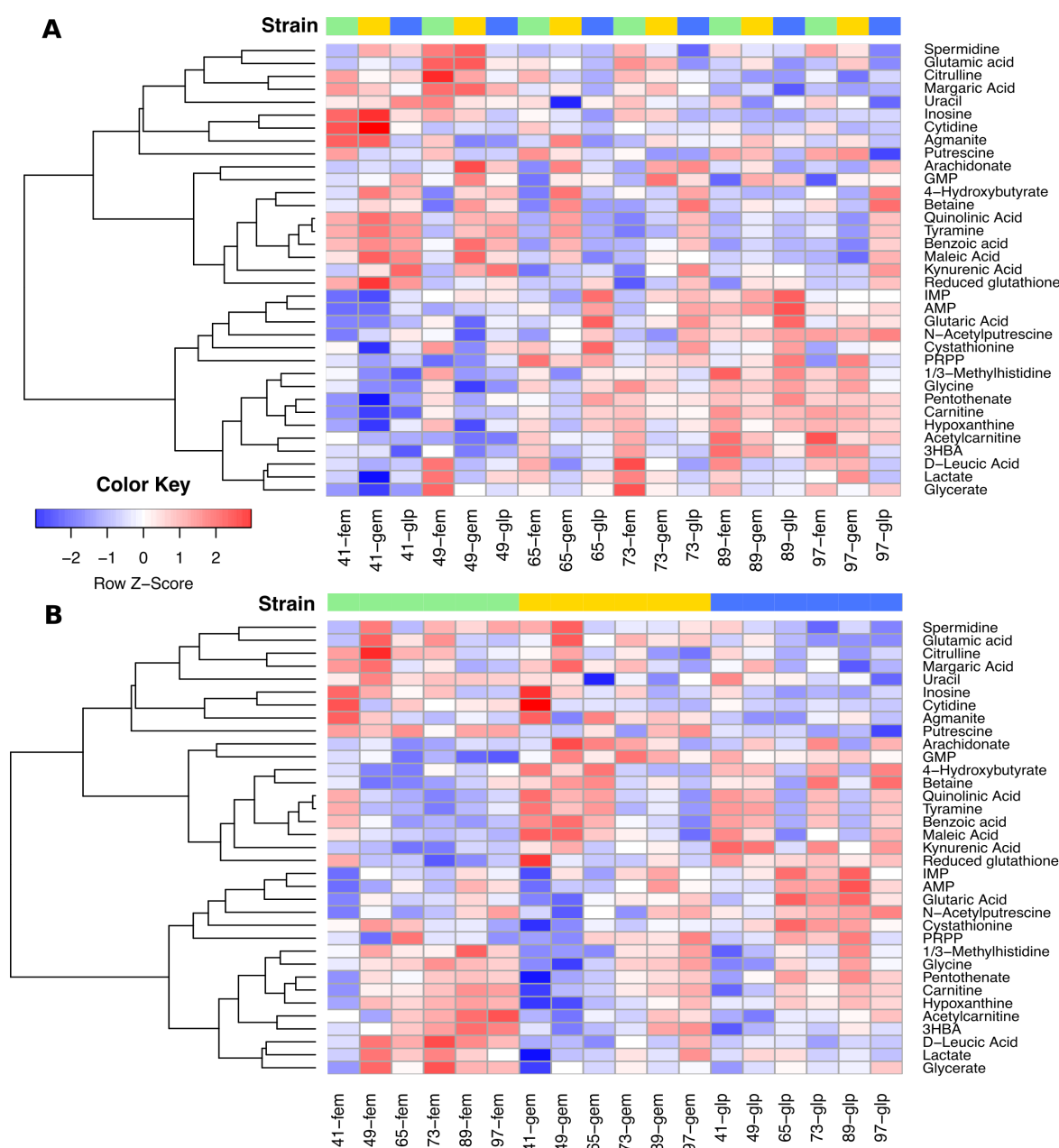


Figure 2.11 Heat map of metabolites that differ by strain

The heat map shows all the metabolites that are significantly differing by strain in the early ageing dataset with the columns ordered by (A) age and then strain, or (B) strain and then age. Strains are indicated by column side bar colours: blue for GLP, green for FEM and gold for GEM.

loss of stress resistance in the FEM and GEM strains but not the GLP. These metabolites are also prominently involved in energy metabolism, so we also investigated strain and age-related differences in some additional markers of energy metabolism.

The ratio of ATP to ADP is a commonly used marker for energy metabolism. Unfortunately, in our dataset ATP was not detected at all, and ADP was removed during data

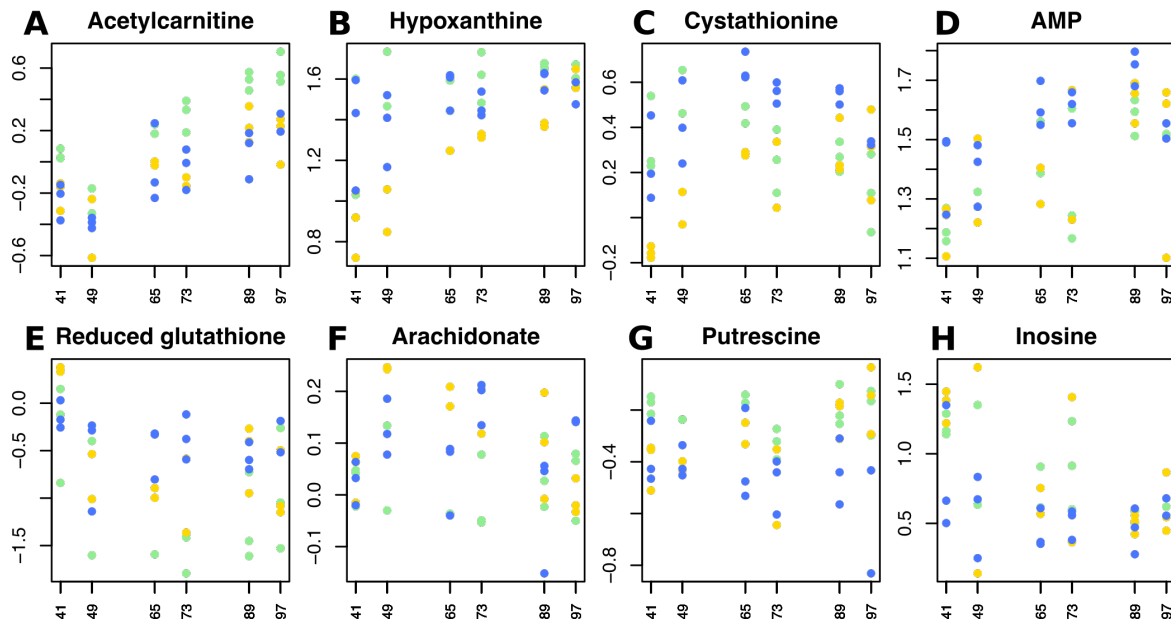


Figure 2.12 Selected metabolites showing characteristic between-strain differences.

Each plot represents a single metabolite's scaled intensity values during early ageing, by time point and strain. Points represent the means of available replicates at that time point and strain. Strain is indicated by colour: blue for GLP, green for FEM and gold for GEM.

pre-treatment due to a large percentage of missing values (Figure 2.2A). While measurements of AMP were available, they are perhaps less informative on their own. AMP was neither age-associated nor significantly associated with time of day, but it does show an effect of strain and is slightly higher on average in GLP than FEM/GEM (Figure 2.12D).

Another estimate of energy availability is the ratio of NAD⁺ to NADH. We did not directly measure these cofactors, but the ratio of pyruvate to lactate can be taken as an estimate as the related enzymes sense shifting levels. Copes et al. (2015) observed a 3-fold decrease in NAD⁺ levels and 20% increase in NADH levels, leading to a 3-fold decrease in NAD⁺/NADH ratio with age, which they were able to estimate by a similar decrease in the pyruvate to lactate ratio. However, we do not see a consistent decrease in this ratio in our dataset (Figure 2.13). The levels of both pyruvate and lactate seem to increase in the oldest time point, and so does the ratio (indicated as "PYR/LACT" in Figure 2.13).

Copes et al. (2015) also examined the relative ratios of NADP⁺ and NADPH as estimated using the ratio of pyruvate to malate, which in our dataset is similarly inconclusive ("PYR/MAL" in Figure 2.13).

The ratio of reduced glutathione to oxidised glutathione indicates redox state. Reduced glutathione should be substantially more abundant than oxidized glutathione, but this ratio is reduced in neurodegenerative diseases and other pathological conditions (Caito

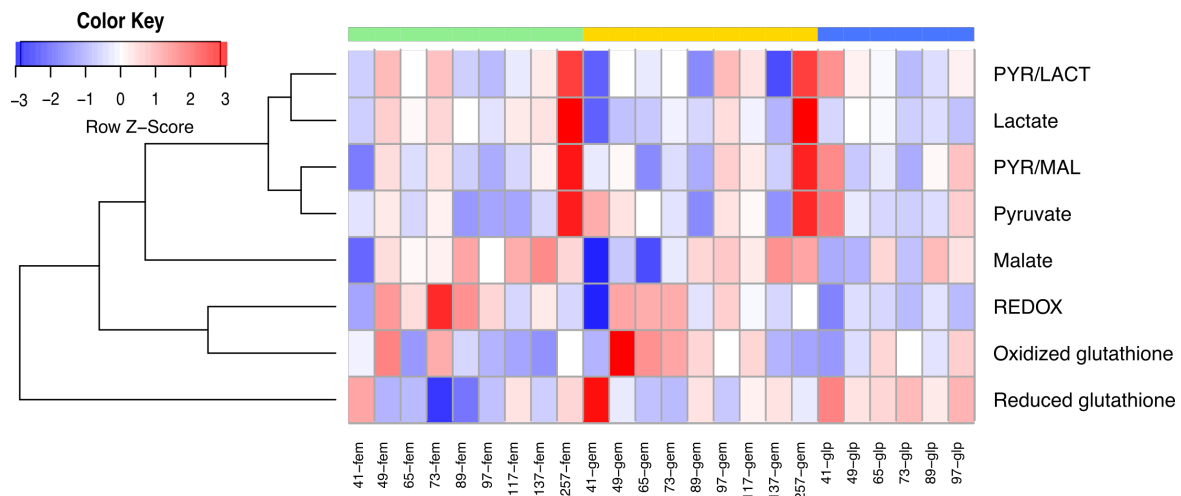


Figure 2.13 Availability of energy metabolites.

This Figure shows a heat map of energy metabolites, antioxidants and their ratios as estimates of energy availability and redox status. Rows are clustered algorithmically, while columns are ordered first by strain and then by age. Strain is indicated by column bar colour: blue for GLP, green for FEM and gold for GEM.

and Aschner, 2015) and is expected to decrease with age (Back et al., 2012). We do see a suggestive if not very clean decrease with age in our dataset (“REDOX” in Figure 2.13).

Finally, there are several examples of sub-clusters with differing and non-linear trajectories with respect to time. For example, putrescine (Figure 2.12G), a polyamine representing degradation pathways and (as discussed above) implicated in *C. elegans* ageing, becomes depleted in GLP but not the other two strains later in ageing. As this metabolite strongly accumulates in later ageing (Figure 2.9A and G, cluster 6), the depletion at this time point in GLP can be taken as indicative of an anti-ageing trajectory, which stands alongside the aforementioned evidence that GLP has a better redox state due to higher levels of reduced glutathione (Figure 2.12E) to suggest a more metabolically youthful profile extends for longer in these long-lived mutants.

Furthermore, arachidonate (Figure 2.12F), an ω -6 polyunsaturated fatty acid (PUFA) illustrates non-linear patterns with respect to time that are noticeably different across all three strains. Arachidonic acid is one of several PUFAs that have been found to be beneficial to *C. elegans* lifespan through pathways that promote starvation resistance and longevity through activating autophagy (O’Rourke et al., 2013). The non-linearity of these complex trajectories during early ageing points at a reason that some of these important longevity-modulating metabolites, such as arachidonic acid, were not revealed as age-associated in the initial analysis of the full ageing time series. One possible explanation for some of the metabolic non-linearity will be explored in the next section.

2.3.7 Some metabolites vary with time of day, in what may reflect circadian metabolic oscillations

The availability of more than one sample per day during early ageing allows us to examine more complex temporal dynamics in metabolism than single-time-point-per-day sampling allows. Moreover, as has been observed in the earlier sections of this chapter, some of these temporal dynamics are clearly non-linear, yet regular with respect to sample collection age. One of the patterns that has been visually apparent for some metabolites is a hint of periodicity, or repeated shifts up and down between neighbouring time points, during early ageing. For example, cluster 3 in Figure 2.9 appears ‘stripey’ in the early ageing time points, with recurring level shifts within each day.

These temporal fluctuations during early ageing are consistent with the possibility of a metabolic response to circadian or other biorhythms. Circadian oscillations in metabolic intermediates have been shown to be widespread and essential for health in mice (Dyar et al., 2018), and it is known from epidemiological studies in humans that deregulated circadian clocks lead to a greater burden of metabolic diseases and cancer in e.g. shift workers (Kecklund and Axelsson, 2016). In *C. elegans*, circadian rhythms with metabolic effects are known to occur despite the fact that in the wild, nematodes live underground (Migliori et al., 2011; van der Linden et al., 2010). They have been shown to be responsive not only to the presence and absence of light, but also to fluctuating environmental temperatures (Goya et al., 2016), although these factors are less clear in laboratory conditions. The metabolic impact of these fluctuations may be partly mediated by feeding behaviour, but not entirely: the same rhythms have been linked to concomitant changes in the ability of the organism to mount a response to stress (Simonetta et al., 2008), and responses to different types of stressor (osmotic vs. oxidative) vary in anti-phases, thus are not all aligned with differences in feeding behaviour. A 12-hour cell autonomous oscillation in ER and mitochondrial genes impacting stress responses has also been observed to be conserved from mammals to *C. elegans*, controlled by the master regulator of the unfolded protein response, *xbp-1* (Zhu et al., 2017). The nature of the relationships between circadian rhythms and downstream metabolite concentrations during ageing are yet to be explored in detail.

Thus, we decided to investigate whether time of day was a significant determinant of metabolite levels in our early ageing dataset. With only two time points per day, we cannot really determine if true circadian oscillations are present, but we can check if level fluctuations track the time of day of sample collection. Samples were collected at the beginning of the day (morning) and at the end of the day (evening) as follows:

- Day 1: 41 h → morning, 49 h → evening;
- Day 2: 65 h → morning, 73 h → evening;
- Day 3: 89 h → morning, 97 h → evening.

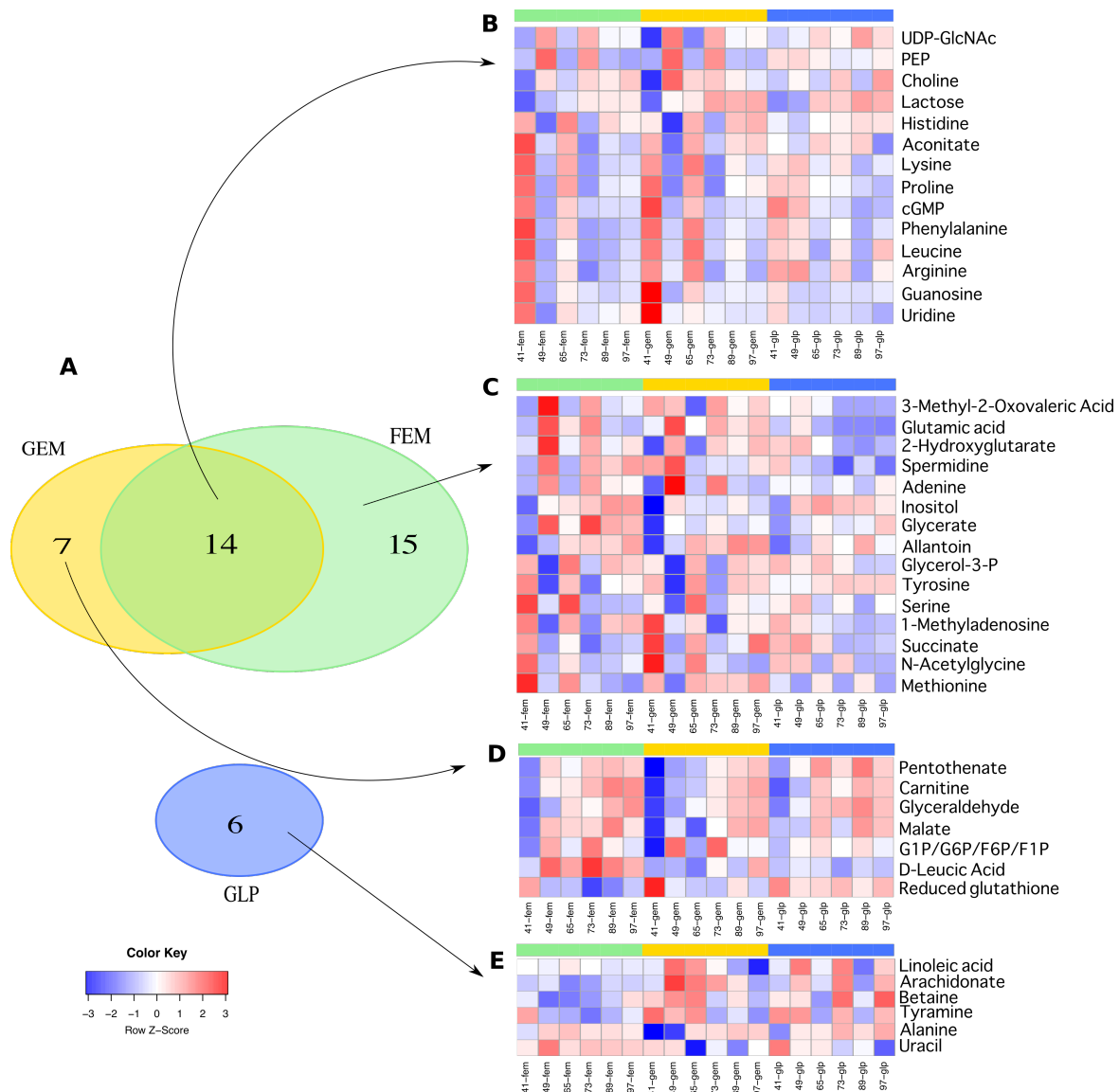


Figure 2.14 Time of day effects in early ageing metabolism.

(A) Venn diagram and (B-E) associated heat maps for metabolites in each strain and group showing time of day effects during early ageing. The Venn diagram (A) indicates the overall numbers of metabolites varying in this way that are unique to each strain and shared between the strains, while the heat maps (B-E) show the individual metabolite profiles for each of these groups. Strain is indicated by colour: blue for GLP, green for FEM and gold for GEM.

As described in the Methods, we used analysis of variance (ANOVA) to determine which metabolites fluctuated with time of day significantly after accounting for the fluctuations due to age in general. There are significant strain differences with respect to these fluctuations, thus, we performed this analysis on a per-strain basis. As illustrated in Figure 2.14, using a threshold of 0.01 for the statistical significance of the time of day effect, we were able to distinguish 42 metabolites with time of day effects in at least one strain. Each strain

has a distinct profile of such metabolites, but GEM and FEM also share a significant subset, while GLP has a different profile of time-of-day varying metabolites altogether.

Taking the metabolites that were varying with time of day in a shared manner between GEM and FEM (Figure 2.14B), we find a substantial cluster of these shared metabolites are more abundant in the mornings and lower in the evenings; these metabolites include proteinogenic amino acids histidine, lysine, proline, phenylalanine, leucine, and arginine, and the nucleoside guanosine and nucleotide uridine. More amino acids are apparent in the FEM-only group (Figure 2.14C), namely tyrosine, serine and methionine, and visual inspection shows that the trend for these metabolites is actually shared with GEM even though their periodicity may not have reached statistical significance in the GEM strain.

It is interesting to speculate why we might be seeing higher levels of amino acids in the mornings in our two wild-type-like strains FEM and GEM. Recently, a comprehensive atlas of mouse tissue-specific 2-hourly metabolomics revealed that more than 50% of the metabolome exhibited circadian temporal oscillations (Dyar et al., 2018), attributing amino acid temporal oscillations to shifts in feeding and exercise behaviour. Feeding and locomotory behaviour may well be the reason for the shifts between high morning and low evening levels of amino acids in *C. elegans* as well. Although we did not track behavioural phenotypes in our study, locomotory behaviour in *C. elegans* has been shown to peak around the early morning in worms (Simonetta et al., 2009). Thus, a prominent hypothesis arising from this data would be that rates of amino acid concentrations are higher when worms are more active, in terms of both moving and feeding.

There is a smaller cluster of metabolites in the shared GEM and FEM group that are depleted in the mornings and higher in the evenings. These are UDP-GlcNAc (uridine diphosphate N-acetylglucosamine, a nucleotide sugar), PEP (phosphoenolpyruvate, involved in glycolysis and gluconeogenesis), choline (an essential nutrient and component of lipid head groups) and lactose (another sugar). Looking at the FEM-only group (Figure 2.14C) and the GEM-only group (Figure 2.14D), we can add further metabolites that are depleted in mornings and higher in the evenings, including spermidine (a polyamine), adenine (a purine nucleobase), glycerate (a sugar obtained from oxidation of glycerol), pantothenate (an essential dietary vitamin B5), carnitine (energy metabolite involved in mitochondrial fatty acid oxidation), glyceraldehyde (another sugar), malate (a TCA cycle intermediary) and glucose-1/6-phosphate (involved in glycolysis).

It is very interesting to note that the time-of-day-varying metabolites in GLP (Figure 2.14E) have a different profile from GEM and FEM altogether. Depleted in the morning and more abundant in the evenings are fatty acids linoleic acid and arachidonic acid (arachidonate), as well as betaine, tyramine and alanine. Uracil, by contrast, is elevated in the mornings and depleted in the evenings. It is well known that the GLP animals have a radically different profile of fats compared to wild type animals (Amrit et al., 2016;

O'Rourke et al., 2009; Steinbaugh et al., 2015). This altered fatty profile has been linked to altered transcriptional programming that in turn prolongs lifespan in germline-free animals. Linoleic acid and arachidonic acid are two polyunsaturated ω -six fatty acids, which are known to have strongly beneficial health effects. Moreover, in the full clustered early ageing heat map (Supplementary Figure A.5), it is apparent that there are other fatty acids and lipids clustering with these key metabolites and fluctuating similarly in GLP and not in GEM or FEM (although not to the same extent of meeting our threshold for statistical significance). For example, citraconic acid, maleic acid, 12-HETE (12-hydroxyeicosatetraenoic acid, a derivative of arachidonic acid) and quinolinic acid (a neurotoxin that increases ROS when supplemented to *C. elegans*, (da Silveira et al., 2018)), kynurenic acid and reduced glutathione all form a part of the same cluster as linoleic acid and arachidonic acid, while 4-hydroxybutyrate (also known as gamma-hydroxybutyrate or GHB, a central nervous system depressant that has been linked with lethargus in *C. elegans* (Dabbish and Raizen, 2011), a driver of metabolic changes during development) clusters with betaine.

Assuming feeding behaviour is also increased in GLP during the mornings, this suggests that the organism is shifting between a daily phase in which stores of these health-positive fatty acids and storage metabolites are built up, and a nightly phase in which stores are reduced. In support of this hypothesis, there is a (not statistically significant) trend towards an anti-correlation with TCA cycle metabolites malate and aconitate, at increased levels when the fatty acids are at lower levels and at lower levels when the fatty acids are at higher levels (Supplementary Figure A.5), suggesting that there is a day/night switch between programs for energy storage and energy metabolism.

Moreover, intriguingly, while on average the FEM and GEM level shifts appear more apparent on days 1 and 2 than on day 3, the GLP level shifts are more apparent on days 2 and 3 but are largely absent on day 1. This hints at a metabolic shift taking place in GLP animals during day 1, which is consistent with large-scale transcriptional changes described in the associated dataset (Mains, 2018). We might expect that circadian oscillations in wild type *C. elegans* might be dampened with age, consistent with the reduced magnitude shifts observed by day 3 in GEM and FEM, as Goya et al. (2016) observe that circadian oscillations die off after day 4 of adulthood. They hypothesise that this is due to the loss of synchronization between individual worms in populations, but another possibility is that it is due to an ageing effect that is not explored in their study, consistent with the onset of early age-related phenotypes noted by Ezcurra et al. (2018) and Ben-Zvi et al. (2009). In this context, it would also make sense that the GLP level shifts carry on and even appear to strengthen into day 3, although of course it is impossible to draw any robust conclusion when the metabolites involved are so different.

It is intriguing to wonder how much of the time of day variance in the metabolome is directly caused by feeding behaviour as opposed to controlled by transcriptional changes

in gene expression that are themselves determined by clock genes or related oscillatory programs. Gene expression oscillates extensively during *C. elegans* larval development (Hendriks et al., 2014), driving the four larval stages with intervening periods of lethargus. It is to be expected that the metabolome would be affected by these massive waves of synchronised transcriptional oscillation. We might wonder if any of the time-of-day variation we are observing in our post-developmental dataset might be due to continued run-ons of developmental oscillations in gene expression. However, consistent with the literature we have observed in the linked transcriptomics dataset that the relevant genes no longer oscillate at our time points (data not shown).

Zhu et al. (2017) report that in *C. elegans*, *xbp-1* controls a cell-autonomous 12-hour cycle that increases stress resistance, and when this cycle is lost, stress resistance decreased. In our linked transcriptomics dataset, we can observe that *xbp-1* does appear to be rhythmically shifting temporally, although our temporal density does not allow us to distinguish 12 hour cycles from 24 hour cycles, in GLP and possibly (after accounting for a general dampening due to day) in GEM, while it is generally lowly expressed in FEM (Figure 2.15). When compared to the metabolomics dataset, varying levels of *xbp-1* are *in phase* with the amino acid levels, in that the levels are higher in the morning time point and lower in the evening time point, and thus *out of phase* with the shifting fatty acid levels in GLP which are lower in the mornings. Taken together, this may provide further evidence for a complex interrelationship between fats, stress resistance and longevity in *C. elegans*.

2.4 Conclusion

It is apparent from this longitudinal metabolomics dataset that many synchronised and unsynchronised, sharp and gradual shifts in metabolism take place during the course of adult life in *C. elegans*. We observed that temporal factors drive variability in the metabolome both in advanced age and, in a non-linear fashion, early in ageing. Several metabolites were determined to be significantly changing with age in a coordinated fashion, including those that become depleted with age, such as oxaloacetate, glutamine and guanosine, and those that accumulate with age, such as 3-hydroxybutyric acid, allantoin and cadaverine. There are clear metabolic shifts during the course of aging, with a prominent decrease in the levels of amino acids, while their polyamine derivatives increase. In addition, many metabolites involved in the TCA cycle show a decrease in level. In general, many of the changes we observed concurred with published literature, but we saw that for some metabolites, the fold change when comparing young to old were dependent on the exact time of day that the samples were taken, tying in to the observed periodicity during early ageing in a significant portion of the metabolome.

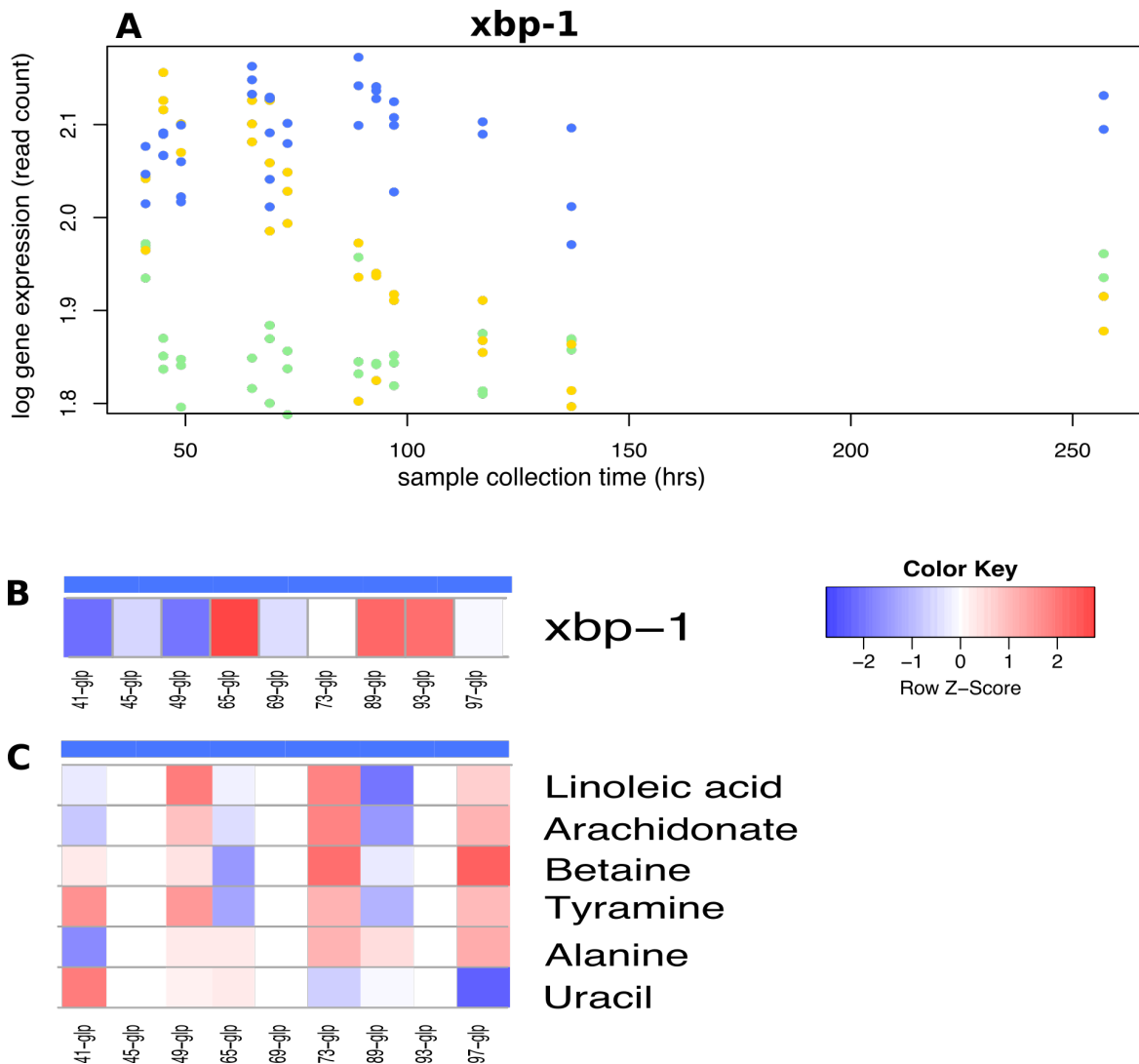


Figure 2.15 *xbp-1* as one candidate transcriptional mediator of time of day variation in metabolite concentrations.

The Figure shows: (A) the normalised log gene expression (read count) values for each sample for the gene *xbp-1*, illustrating that for the GLP strain, some within-day temporal shifting is apparent for this gene during the early ageing time points; (B) a heat map of the mean GLP *xbp-1* expression values, and (C) a heat map of the time-varying metabolite concentration levels for those metabolites that show time-of-day variance in the GLP strain. Strains are indicated by colour: blue for GLP, green for FEM and gold for GEM.

In terms of pathways, we saw that amino acids were significantly enriched for age-associated metabolites, again in keeping with previous literature. Many amino acids are also known to be longevity modulators, and we observed a general correlation between the groups of metabolites that were age-associated and those that were longevity modulators. It is interesting that amino acids were also among those metabolites that vary in early ageing in GEM and FEM in a time-dependent manner, hinting at cycles of nutrient depletion and

replenishment. For some of the amino acids, the age-associated phenotype of depletion is similar to the smaller scale shifts in metabolism that take place on a daily basis.

We did not observe a clean separation between the strains in the metabolome, and only more subtle changes are apparent between the strains, particularly the long-lived GLP strain as compared to the FEM and GEM strains. The most striking strain difference observed is in the exact temporal dynamics of what appeared to be periodicity in the early ageing samples, with amino acids and sugars peaking at different times of day in GEM and FEM hinting at possible shifts in feeding behaviour, while in GLP temporal shifts are apparent mainly in lipids and fatty acids. Comparisons of these temporal shifts in the metabolome across these three non-reproductive strains reveal that distinct metabolic ‘programs’, patterns or profiles may be activated at different times in the different genetic backgrounds. But based only on this observational dataset, it is not possible to determine which of these changes necessarily go together, and which are merely contingently associated.

A more powerful *in silico* approach would ideally be able to make use of biochemical knowledge about how different metabolites relate to each other within the overall metabolism of the organism, in order to make mechanistic predictions relating shifts in metabolism to overall phenotypes. This is the objective of whole-genome metabolic modelling, and it is to this approach that we will now turn. As a prerequisite for such approaches, we will need to use a whole-genome model of the metabolism of *C. elegans*, which is the topic of the next chapter.

3. A consensus genome-scale model of *C. elegans* metabolism

3.1 Introduction

A genome-scale metabolic reconstruction model (GSMM) is a computational representation of the full range of known metabolic pathways and reactions across an entire genome, with associated annotations of genes, for a given organism (Thiele and Palsson, 2010). Such models have long been in use for unicellular micro-organisms, but have only recently begun to be applied in more complex organisms.

Efforts to reconstruct the metabolism of *C. elegans* culminated in 2016 with the back-to-back publication of two curated and validated whole-genome metabolic models (Gebauer et al., 2016; Yilmaz and Walhout, 2016). This was followed by a third published model shortly thereafter (Ma et al., 2017).

Each of these efforts used different strategies to reconcile their metabolic knowledge bases, and were represented using different annotation standards, thus, they were difficult to mutually reconcile. Each had different strengths and weaknesses, and while they naturally overlapped to a large extent, they were not completely overlapping – each model had some unique aspects, as well as different choices of representational abstraction that hindered a simple combination of their contents. Thus, it was desirable to obtain a merged and unified model which contained the best of all the knowledge represented in each of the individual published models, while resolving duplication across them.

With this in mind, our group harnessed an opportunity offered by GENiE (an EU COST action, www.worm-genie.eu) to host a workshop locally at the Babraham Institute to initiate a community effort for reconciling and extending a consensus model of *C. elegans* metabolism. The workshop was held in April 2017 and was a great success (Hastings et al., 2017). The community that was established was christened ‘WormJam’ (short for *Worm Jamboree*) and has since resulted in the creation and extension of a consensus model and development of infrastructure to support the ongoing curation and extension thereof.

In this chapter, the development of and current status of this consensus model is described.

3.2 Methods

3.2.1 Whole-genome metabolic models

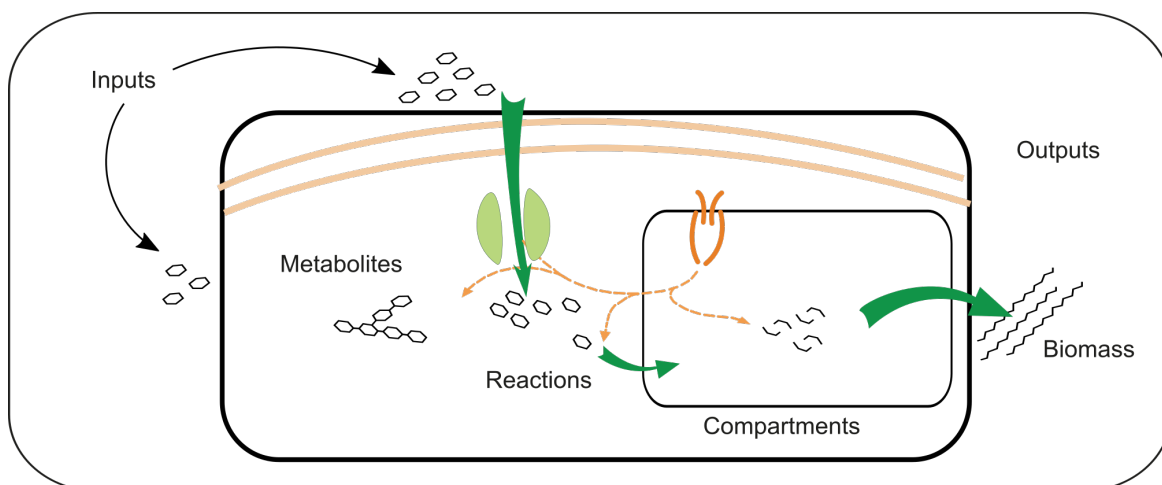


Figure 3.1 A schematic illustrating the structure of whole-genome metabolic models.

Whole-genome metabolic models describe all the metabolic reactions taking place in an organism abstracted to the level of a single cell. Reactions take place in compartments which represent cellular organelles, such as the cytosol, mitochondria and nucleus. The system as a whole accepts inputs from the extracellular environment and creates outputs which are transported out of the cellular environment.

Whole-genome models describe intracellular metabolism as a network of reactions between metabolites (Thiele and Palsson, 2010). The general structure of these models is illustrated in Figure 3.1. Each metabolite that is produced, consumed or intermediary in any metabolic reaction is represented, including simple amino acids such as L-alanine, metal ions such as Cu^{2+} , and complex cofactors such as NAD^+ and CoA. Metabolites are related as inputs and outputs in biochemical reactions, which are stoichiometrically balanced. Reactions take place in compartments, which represent cellular locations, such as cytosol, mitochondria and nucleus. The system as a whole is bounded, and specified input and output reactions are allowed at the system boundary to the extracellular environment, representing for example nutrient and oxygen uptake and metabolic byproduct excretion.

Reactions are further associated with logical rules specifying which gene products or combinations of gene products are known to catalyse those reactions. Depending on their biochemical constraints, reactions are potentially bidirectional, carrying either positive or negative flux (reaction throughput) depending on the overall required flow of metabolites.

Reactions are furthermore associated with rules for the maximum and minimum flux that they can carry, which are ideally based on experimental knowledge. Although for the vast majority of reactions in the knowledge base the true experimental maxima and minima are not known, constraining key exchange reactions for nutrients and oxygen uptake

typically limits the flux through the remainder of the network to realistic levels. A final special type of reaction is the biomass reaction, which specifies the metabolic components of the structure of the organism (e.g. nucleic acids, proteins and lipids), in the relative stoichiometric quantities in which they are typically required in the growing organism to create a unit of biomass, scaled so that the flux through this reaction equals the growth rate μ of the organism.

For *E. coli*, a community effort in annotation of the whole-genome metabolism has resulted in a widely used model for strain K-12 MG1655 (Orth et al., 2011) and several other models for alternative strains and conditions. Similar reconstructions exist for human cells (Swainston et al., 2016; Thiele et al., 2013) and in particular for core central human metabolic pathways (Smith et al., 2017). The typical size of such models is thousands of reactions and metabolites, across several compartments. These models have become an essential tool for the study of metabolism in a wide range of cellular functions.

3.2.2 The WormJam model annotation and development pipeline

A custom pipeline for the development and extension of the shared consensus model was developed using SBTAB (Lubitz et al., 2016), a format for representing model contents as spreadsheets. The model is represented as SBTAB and the resulting spreadsheets – for reactions, metabolites, genes, and pathways – are shared as Google spreadsheets in a folder to which all members of the WormJam community have access via shared membership of a WormJam Google group. Through these shared spreadsheets, all members of WormJam are able to contribute to model curation.

The use of standards and shared identifiers in the representation of such models greatly aids in their use and adoption by the scientific community, as well as facilitating the community-wide integration of different models to achieve consensus reconstructions (Herrgård et al., 2008). The standard representational format for models is the Systems Biology Markup Language (SBML; Hucka et al., 2003). Individual entities within the model may be annotated with meta-data cross-references to various databases and resources that enable disambiguation of the entity, such as ChEBI for metabolites (Degtyarenko et al., 2008; Hastings et al., 2016) or UniProt for gene products (Bairoch et al., 2005).

Periodically, the contents of the spreadsheet are built into a working model SBML file which is then validated before being released. The build process uses a custom Java script, which is available from the online WormJam repository¹. Releases and other project updates are advertised by email to the WormJam Google Group².

¹ Website: <http://gh.wormjam.life/>

² Email: wormjam@googlegroups.com

3.2.3 Model validation

Genome-scale metabolic models can be validated along at least two distinct dimensions: how *complete* the model is relative to the best known biochemistry, and how *correct* the model is. Completeness can be benchmarked for example against databases and other models of the same organism, while correctness can be checked manually but also tested with quantitative simulations.

To validate the correctness of the model quantitatively, part of what is commonly done is to check the model's ability to make predictions. In general, the *in silico* predictions made by such a model, which relate to growth under different input conditions as well as differential use of pathways within the overall reaction network, need to be compared to what has been measured *in vivo*.

Growth and energy metabolism

The most basic validation that can be applied to a genome-scale model is *can it simulate growth?* To do this the model is used together with Flux Balance Analysis, a mathematical simulation method that is able to make predictions about growth given various allowed inputs, which will be described in full in the subsequent chapter.

In the validation steps performed for the WormJam model, the most primary validation for each release was whether the model could 'grow', or carry flux in an FBA simulation, implemented in Python using the CobraPy library (Ebrahim et al., 2013) (substantial details presented in the subsequent chapter). Blocked reactions were assessed using CobraPy's analysis method '`cobra.flux_analysis.find_blocked_reactions`'. Additional information was sometimes given by the method '`cobra.flux_analysis.reaction.assess`', which tests for a given reaction which of the inputs cannot be acquired from the model under current conditions and which of the outputs cannot be consumed. However, if the broken chain of fluxes is very long then this method is not able to provide accurate estimations of the problem metabolites.

Furthermore, we evaluated the predicted maximal growth under different scenarios of available inputs in order to benchmark the use of different strategies for energy metabolism in the model. We tested a range of different allowed inputs of food intake (bacterial digestion and glucose) and oxygen consumption. These simulations were also performed in Python with CobraPy.

Essential and bottlenecked gene predictions

The ability of whole-genome models to predict which genes are essential for growth or active metabolism, i.e. predicted gene essentiality, is considered an important metric for the quality of the annotations in a given model.

Essential genes are those that if knocked out (i.e. flux set to zero for reactions that have that gene annotated as enzyme), the model cannot grow at all. Model-predicted essential genes are typically compared to genes that are known to be lethal if knocked out. Of course, there may be many reasons that a gene that is known to be lethal if knocked out may not be predicted to be essential for sustaining growth in the model, and vice versa. Metabolic models necessarily exclude all the complexities of cellular signalling, transcription and translation. Moreover, tissue specific or sub-cellular location-specific gene activity is mostly not annotated in the model, and even in cases where a particular reaction may be localised to e.g. cytosol or mitochondria, it is often not 100% known which genes are differentially needed in the different compartments. Thus, model annotations contain redundancy in their gene annotations. Moreover, many of the model pathways contain reactions that are predicted, or that are needed to plug gaps that would otherwise exist in the model if we only included those parts of metabolism for which we have biochemical evidence in the appropriate organism. On the other hand, there are artificial bottlenecks in the model where insufficient alternative pathways have been included in the model.

For the WormJam model validation reported in this chapter, we compared the predicted essential genes in the model to those that are known to be essential for *C. elegans* as per the annotations of lethal phenotypes in the WormBase database, typically based on RNAi screening (Chen et al., 2005; Lee et al., 2018).

3.2.4 Visualisation of arbitrary reaction groups

Typically, for large models of this type incorporating thousands of metabolites and reactions, it is not possible to comprehensively visualise all the content at a level of detail which allows tracking individual metabolites and paths through reaction chains. Various tools exist which allow metabolic network visualisation. For example, SBML models, including genome-scale models, can be loaded into Cytoscape (Su et al., 2014) and the resulting network, in which both reactions and metabolites are nodes, and edges connect reactions with the metabolites that participate in them, can be interactively visualised. Another popular network visualisation tool is Gephi (Bastian et al., 2009). The visualisation result for a generic visualisation tool is rather dependent on the layout algorithm which is used to automatically lay out the content.

Visualisation tools specific for genome-scale models include MetDraw (Jensen and Papin, 2014), MetExplore (Cottret et al., 2018) and Escher (King et al., 2015). While Escher

only allows visualisation of pre-drawn pathway maps and is therefore not useful for whole-model visualisation, MetExplore and MetDraw provide automated layouts. MetExplore provides a rich source of functionality for online editing, visualisation and exploration of genome scale models. However, for reasons that are probably relatively trivial but which I did not investigate further, it failed to parse the WormJam SBML model. MetDraw, in contrast, parses the model and generates a candidate visualisation relatively quickly. The MetDraw approach to automated model visualisation segregates the model content by annotated pathway, and then within each pathway generates a default layout in which metabolites are shown nodes, and reactions are shown as edges connecting metabolites, an intuitive metabolism-oriented simplification on the Cytoscape default which has both reactions and metabolites as nodes. Figure 3.2 shows a small subset of pathways from the WormJam model visualised with the MetDraw tool.

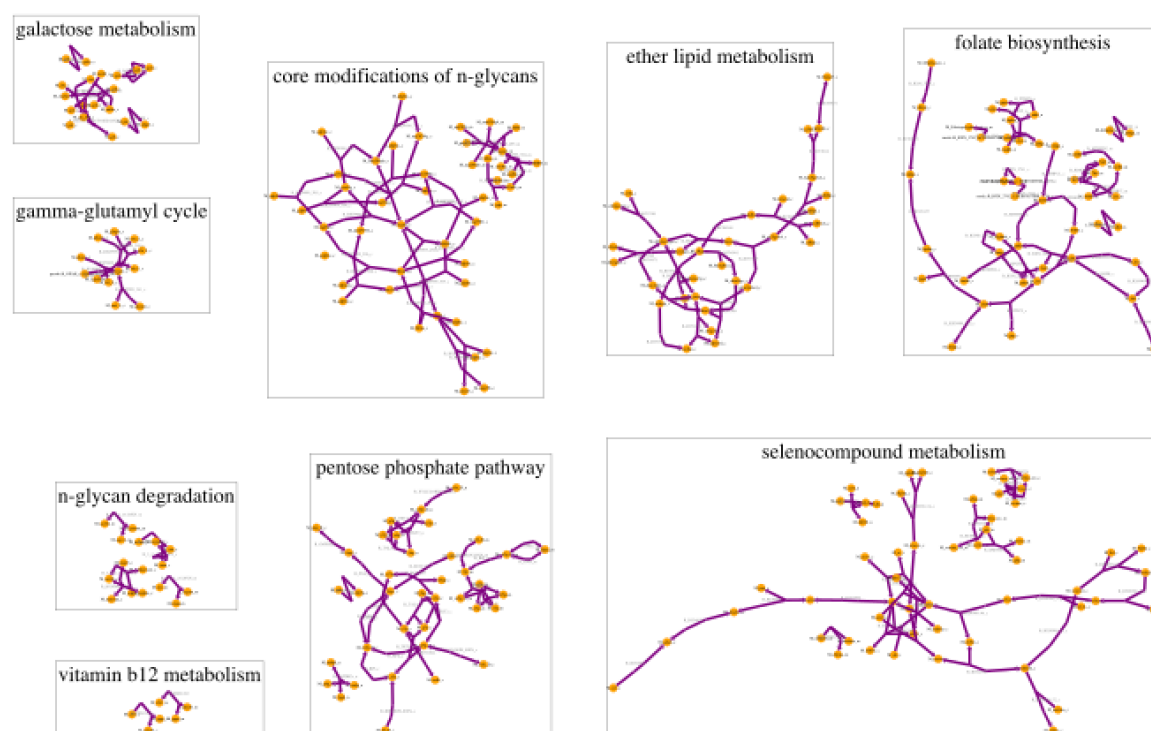


Figure 3.2 A subset of pathways in the WormJam model, visualised with the MetDraw tool. A subset of pathways in the WormJam model selected from the default visualisation of the model created with MetDraw. Within each pathway, metabolites are illustrated as nodes and reactions as edges connecting nodes. The layout is computed automatically.

In addition to segregating the model content by annotated pathway and introducing the simplification of reactions as edges, MetDraw offers another optimisation that is essential for visualising genome-scale models sensibly: it defines ‘minor metabolites’, as those that participate in a very large number of reactions, and does not necessarily connect those up to each other. Without this optimisation, in a network visualisation, a single metabolite

node would be connected to each of the reactions in which it participates, thus joining those reactions together. Minor metabolites represent common cofactors and by-products of metabolism, such as water, hydrogen and ATP. These metabolites appear regularly in reactions that are otherwise completely unrelated, thus it is important to be able to segregate those reactions if they do not share other metabolites. Hence the introduction of non-unified minor metabolites improves the network visualisation.

In what follows I have made occasional use of Cytoscape and of MetDraw for model visualisation. However, I have also implemented my own very rudimentary visualisation library in Python which follows a similar approach to MetDraw for automated layout, but allows arbitrary subsets of selected reactions to be visualised, regardless of their annotated pathways. This was an important tool when I was debugging the various model releases, as it is important to be able to visualise connections between reactions by network rendering when attempting to ascertain blockages which prevent the model from carrying flux. I used the same approach as MetDraw of defining ‘minor’ metabolites to simplify complex reaction chains. Moreover, I added functionality to visualise flux distributions by colour and thickness of reaction edges, which is functionality that is provided in Escher for pre-drawn maps.

3.3 Results and Discussion

3.3.1 Initiation of the WormJam community

At the time of the GENiE workshop held in Babraham in April 2017, it was not yet clear to what extent the previously published *C. elegans* metabolic reconstructions overlapped. The objectives of the workshop were thus:

- To survey the current state of *C. elegans* metabolism research and evaluate the published models;
- To chart a roadmap for how the different metabolic reconstructions, published and non-published, could be integrated in a community-driven effort leading to a comprehensive consensus *C. elegans* metabolic reconstruction;
- To specifically include researchers from both ‘wet-lab’ and ‘dry-lab’ contexts in order to exchange knowledge and experiences, sharing insights into how different types of data can be integrated using metabolic reconstructions.

The workshop was attended by 30 scientists from three continents spanning the full breadth of the planet, showcasing an exciting diversity of cutting-edge *C. elegans* metabolism research. The outcomes of the workshop were published in (Hastings et al., 2017). In summary:

- The overlap, strengths and weaknesses of the existing published models were identified, and several ongoing efforts to mutually reconcile them were identified;
- The gap between metabolites represented in these models and measured in metabolomics experiments was highlighted, and an effort to reduce this gap by working together to fill in holes in the models was prioritized;
- An open community vision for the future was established, which had as an objective the establishment of a community platform for the development and extension of a shared consensus model, and the holding of future annotation workshops.

In the time since then, that WormJam vision became a reality. A shared platform was created online using freely available Google infrastructure, with file sharing using Google Drive and discussions and membership management achieved through use of a Google Group. An initial consensus model was automatically created, and subsequent annotation was enabled using editable spreadsheets powered by the SBTab utility (Lubitz et al., 2016), which allows round-tripping between spreadsheets and SBML models. Several face-to-face and virtual annotation ‘jamborees’ were held, and the resulting model is being used as the basis for further research.

The remainder of this chapter will discuss in further detail several of these steps.

3.3.2 Overlaps and differences between published models

The overlap and differences between the two initially published models – ElegCyc (Gebauer et al., 2016) and iCEL1273 (Yilmaz and Walhout, 2016) – were evaluated along several distinct dimensions.

The numbers of different constituent parts and annotations differed between the models. Table 3.1 lists the relevant indicators.

Table 3.1 Metrics for the size of two published models

Two published reconstructions are shown with indicative metrics for their sizes. The number of unique (across compartments) metabolites is shown in brackets after the total number, as in these types of models, metabolites are distinguished per compartment. Abbreviations: cyt.=cytosol, mit.=mitochondrion, nuc.=nucleus, extra.=extraorganism/extracellular.

| Model | Metabolites (unique) | Reactions | Genes | Pathways | Compartments |
|----------|----------------------|-----------|-------|----------|------------------------------|
| ElegCyc | 2357 (998) | 1921 | 981 | 176 | 4 (cyt., mit., nuc., extra.) |
| iCEL1273 | 1718 (885) | 1985 | 1289 | 98 | 3 (cyt., mit., extra.) |

Using simple name comparisons, I evaluated overlaps in annotations across the models. The results (Figure 3.3) are striking in that the overlaps appear so low, but this is clearly an underestimate of the true overlap that would be obtained with a more comprehensive

reconciliation strategy. Rather, it can be taken as indicative of the plurality of naming schemes for this type of data, and the necessity of using non-semantic standard identifiers for the annotation of content in models of this type (McMurry et al., 2017).

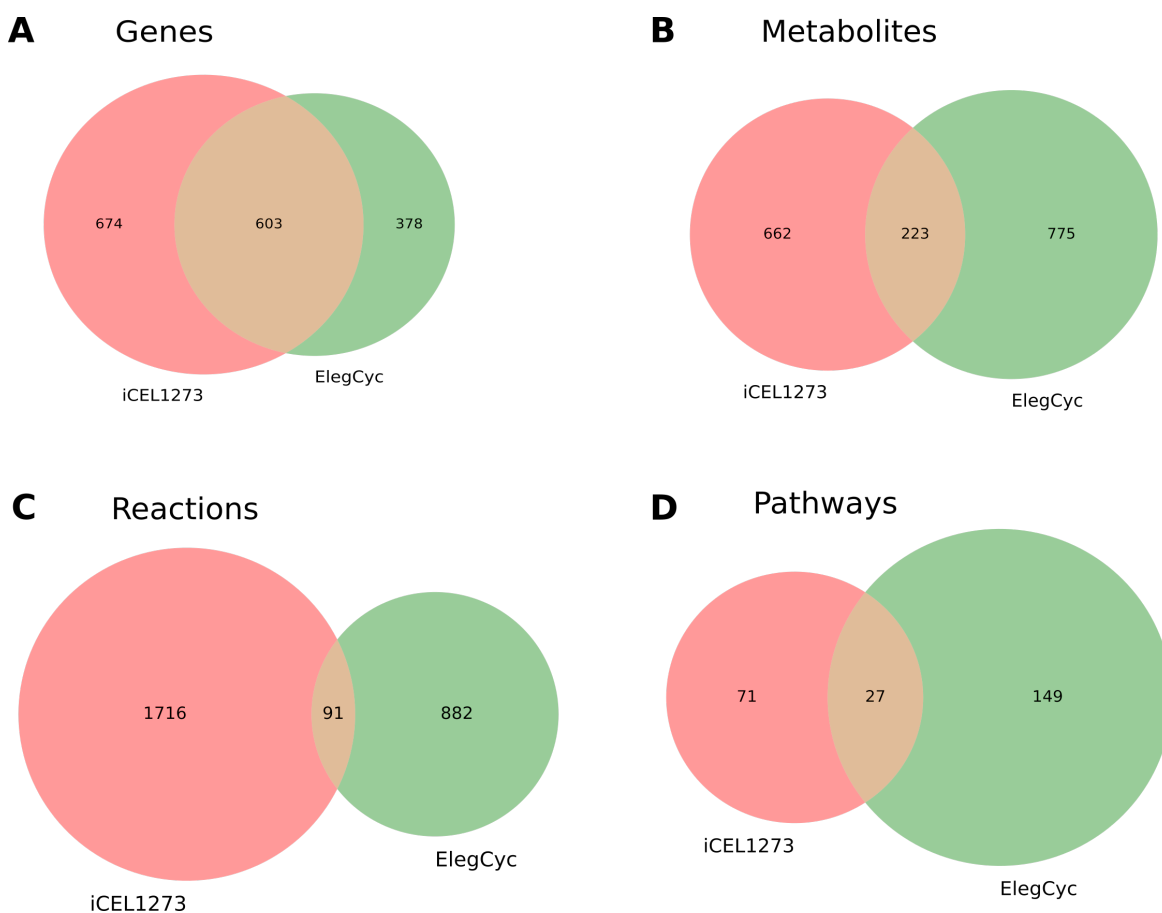


Figure 3.3 Overlaps between two models based on name matching.

Exact, case insensitive name matches show a low level of overlaps between the two models as illustrated for (A) genes, (B) metabolites, (C) reactions and (D) pathways.

An automated merge of these two precursor models was done by two members of the WormJam community – Chintan Joshi and Nathan Lewis – and is described in (Witting et al., 2018). In addition to being partially overlapping, the two models had different strengths and weaknesses. ElegCyc offered the option to choose between an axenic (bacteria-free) and a bacterial food source, and included a representation of the nucleus with details of specific reactions taking place in that compartment. iCEL1273, on the other hand, offered a more detailed representation of *C. elegans* biomass tailored to different stages of the growing organism, and an explicit network of reactions for bacterial digestion showing which aspects of the bacterial food source served as input to which aspects of the worm metabolism. The initial merge retained all unique aspects of both models, but resolved exactly duplicated reactions and metabolites. This meant that the initial build of the merge

had several different options for food intake as well as several different options for biomass. Some of these have been altered or removed in subsequent rounds of manual curation.

Briefly, the steps that were followed in performing the automated initial merge involved merging genes, reactions and metabolites based on their names and annotations. Genes were first merged based on WormBase gene identifiers (which have the form WBGeneXXXXXXXX where XXXXXXXX is a unique number per gene) (Lee et al., 2018). Furthermore, the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017) and the Gene Ontology (Ashburner et al., 2000; The Gene Ontology Consortium, 2019) were queried to obtain additional gene annotations for reactions, to ensure that the gene associations were as comprehensive as possible so that they could be used as input to the merging algorithm.

Metabolites were disambiguated using the BiGG (King et al., 2016), KEGG (Kanehisa et al., 2017), MetaNetX (Moretti et al., 2016) and ChEBI (Hastings et al., 2016) databases. Charge and formula were also extracted for each metabolite. Metabolites lacking any annotated database ID were resolved by manual database searches. Duplicate metabolite entries were merged by removing one of the instances and updating the remainder of the references accordingly.

The most complex merging step was for reactions. The process that was followed checked reaction stoichiometry and gene-reaction associations. Exact duplicates were removed. If one gene-reaction association was a superset of the other, the union of the gene-reaction associations was used. Stoichiometry was also checked without protons and after elemental balancing. However, even after the automated process had completed, many duplicate reactions still remained due to subtle differences in representational choices and differences in gene-reaction association. Many of these were resolved during the subsequent rounds of manual annotation.

3.3.3 Manual annotation of the model

The manual curation effort was completed in several phases or ‘rounds’ of curation.

Infrastructure for manual curation

The infrastructure that was used to support this annotation process consisted of sets of read-only SBTabs spreadsheets for each of the source models, as well as editable SBTabs spreadsheets for the current WormJam model version, as illustrated in Figure 3.4.

In these editable spreadsheets, within each table, content was assigned a unique identifier which was then used for references in other tables. The metabolite table unique identifiers typically followed the form *M_name_compartment* where *name* was, e.g. ‘atp’ for ATP, and *compartment* was a unique code for a relevant compartment, e.g. ‘c’ for

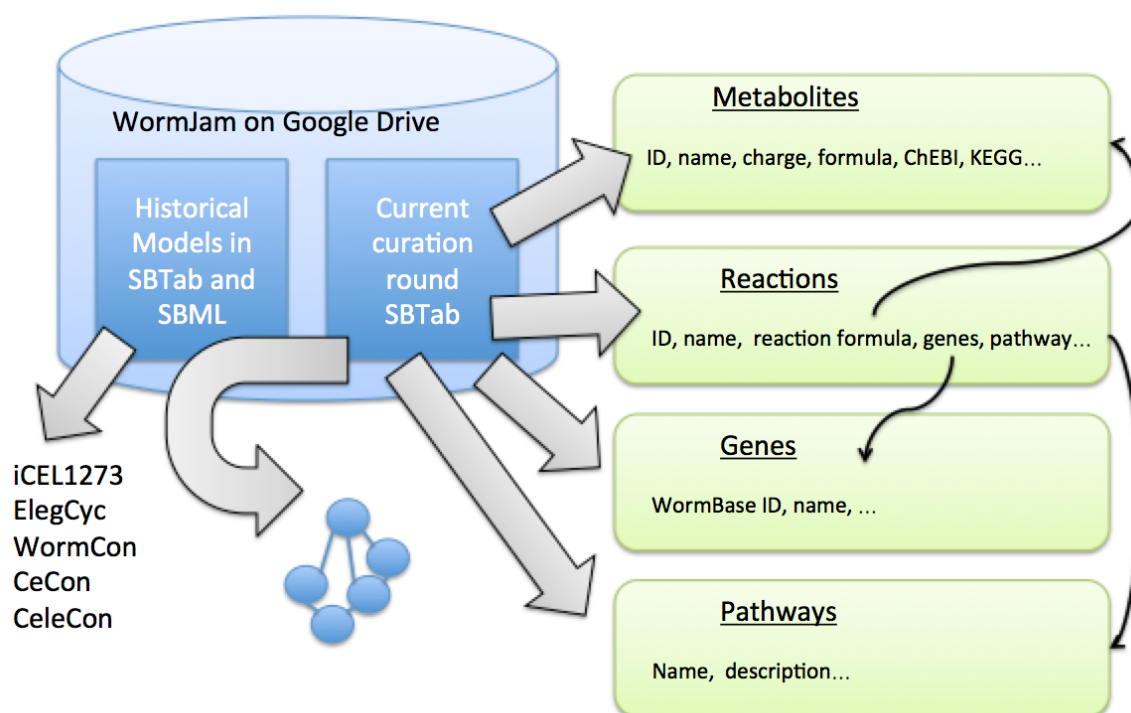


Figure 3.4 Overview of the infrastructure for manual model curation.

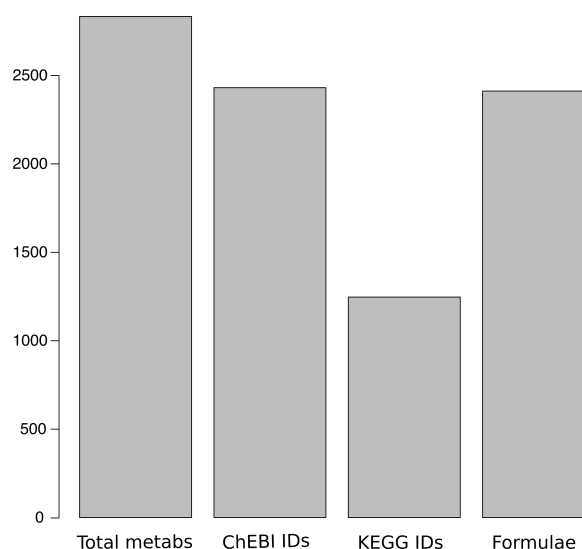
A shared Google Drive folder housed the infrastructure for manual curation. For reference purposes, SBTAB spreadsheets were available for each of the pre-existing *C. elegans* reconstructions. These included iCEL1273 (Yilmaz and Walhout, 2016) and ElegCyc (Gebauer et al., 2016), as well as CeCon (Ma et al., 2017) and the unpublished merges WormCon and CeleCon. Editable spreadsheets included tables for metabolites, reactions, genes and pathways. After each successive round of curation, an SBML version of the model was built and subsequently validated.

cytosol. In the reaction table, identifiers typically have the form R_*rname* where *rname* is a unique compact identifier for the reaction. Genes are primarily identified with WormBase identifiers as mentioned above, while pathways were assigned a simple unique identifying and descriptive name. These identifiers were then used to facilitate references between the tables, for example, metabolites appear in reaction formula descriptions, and genes in reaction gene-product annotations. Automated validations ensure that all metabolites that are referenced in the reaction table also appear in the metabolite table, and so on.

Focus and scope of manual curation rounds

The initial effort of the first round of the community manual annotation process, completed in late 2017, was aimed at the identification and resolution of duplicate reactions that had not been automatically removed (typically because of differences in gene annotation, compartments, or presence of minor cofactors). Moreover, the metabolic literature was

Figure 3.5 Counts of identifiers for metabolites annotated in the WormJam model. These counts are for non-unique metabolites, that is, metabolites localised to compartments within the model. Total metabas=the total number of metabolites in the model. ChEBI IDs=number of annotated ChEBI identifiers. KEGG IDs=number of annotated KEGG identifiers. Formulae=number of annotated chemical formulae.



consulted to verify the presence and localisation of some of the reactions, and some reactions determined to be not taking place in *C. elegans* were removed.

As mentioned, part of the WormJam effort has involved standardised metadata annotation. For metabolites, alongside their within-model identifier and name, charge and formula, ChEBI identifiers (Degtyarenko et al., 2008; Hastings et al., 2016) are annotated as primary external metadata, with KEGG compound identifiers as an alternative (Kanehisa et al., 2017). KEGG reaction identifiers may also be associated with reactions. For genes, WormBase identifiers are the primary metadata annotated (Chen et al., 2005; Lee et al., 2018).

Missing ChEBI (Degtyarenko et al., 2008; Hastings et al., 2016) identifiers were annotated manually for all metabolites in the model. Based on the ChEBI identifiers, mismatches between annotated names, identifiers and charges of the metabolites were corrected. Having ChEBI identifiers associated with the metabolites in the model provided disambiguation and a link to chemical structures, which in turn allows for comparison with metabolomics datasets (Witting et al., 2018). Figure 3.5 shows the numbers of identifiers of each type included in the model. In some cases, this ChEBI annotation process led to the discovery of duplicate metabolites, which often led to identification of a set of associated duplicated reactions as well.

The second community manual curation effort, completed in late 2018, then proceeded to enhance and extend the representation of specific aspects of biology not previously included in any of the merged models. As described in Witting et al. (2018), the representations of glycogen metabolism, fatty acids and sphingolipid metabolism, and epigenetic marks were updated and extended. Pathways, reactions and metabolites were added for *C. elegans*-specific maradolipids and ascarosides. All metabolites present in the model and not yet present in ChEBI were submitted to ChEBI, together with structural representations for

charged and neutral forms. Finally, the list of pathways annotated to reactions in the model was disambiguated, corrected and simplified.

Debugging and validation of each model version

After each phase of the manual curation effort had been completed, the model was rebuilt into computable SBML format from the manually curated spreadsheets in SBTab format for evaluation. At this stage, basic automated checks were completed, such as that all cross-references between reactions and genes and metabolites were present. At this stage, typically after each curation round there were a few such missing references, but these were quick to resolve. The much more difficult part of the validation process was getting each new build to carry flux, which took in each case up to a few weeks.

It was necessary to test that the model could carry flux in all needed pathways in order to create biomass from food sources in FBA simulations. Each build of the model was initially not able to carry flux, hinting at the difficulties involved in manual curation of a model this size. For example, after the first round of curation (late 2017/early 2018), the model contained 1050 reactions that could not carry flux, including the assembly of the primary biomass, while after validation and debugging the release version of the model contained 685 blocked reactions, as described in (Witting et al., 2018).

Each reaction in the model was individually tested with respect to its ability to carry flux, and each metabolite whether it could be produced. It is typical for large-scale metabolic models to include some percentage of reactions that cannot carry flux and metabolites that cannot be synthesized. Also, reactions that are blocked with respect to a given biomass reaction may not be blocked with respect to another, since the block may relate to the take-up of the reaction outputs in the appropriate way on the path to biomass construction. However, it is important that all key pathways are fully functional and biomass can be generated as expected from all cellular precursors.

For each blocked reaction that was on a path towards biomass construction, a step by step ‘debugging’ procedure was completed to determine the broken steps in the various reaction chains between model input and eventual biomass creation. A reaction chain is a series of reactions with the inputs of one being the outputs of another, thus, if one reaction is broken somewhere along the chain, none of the reactions in the chain will be able to carry flux, as everything that is produced must be consumed (or exported). Moreover, of course, reactions do not form simple chains but complex interconnected non-linear patterns (see for example the illustration of a subset of model pathways in Figure 3.2).

Custom visualisation of chains of reactions was very important in arriving at the correct diagnosis of a problem or multiple problems in a reaction chain, as was the information provided by the reaction assessment implementation in CobraPy. For this

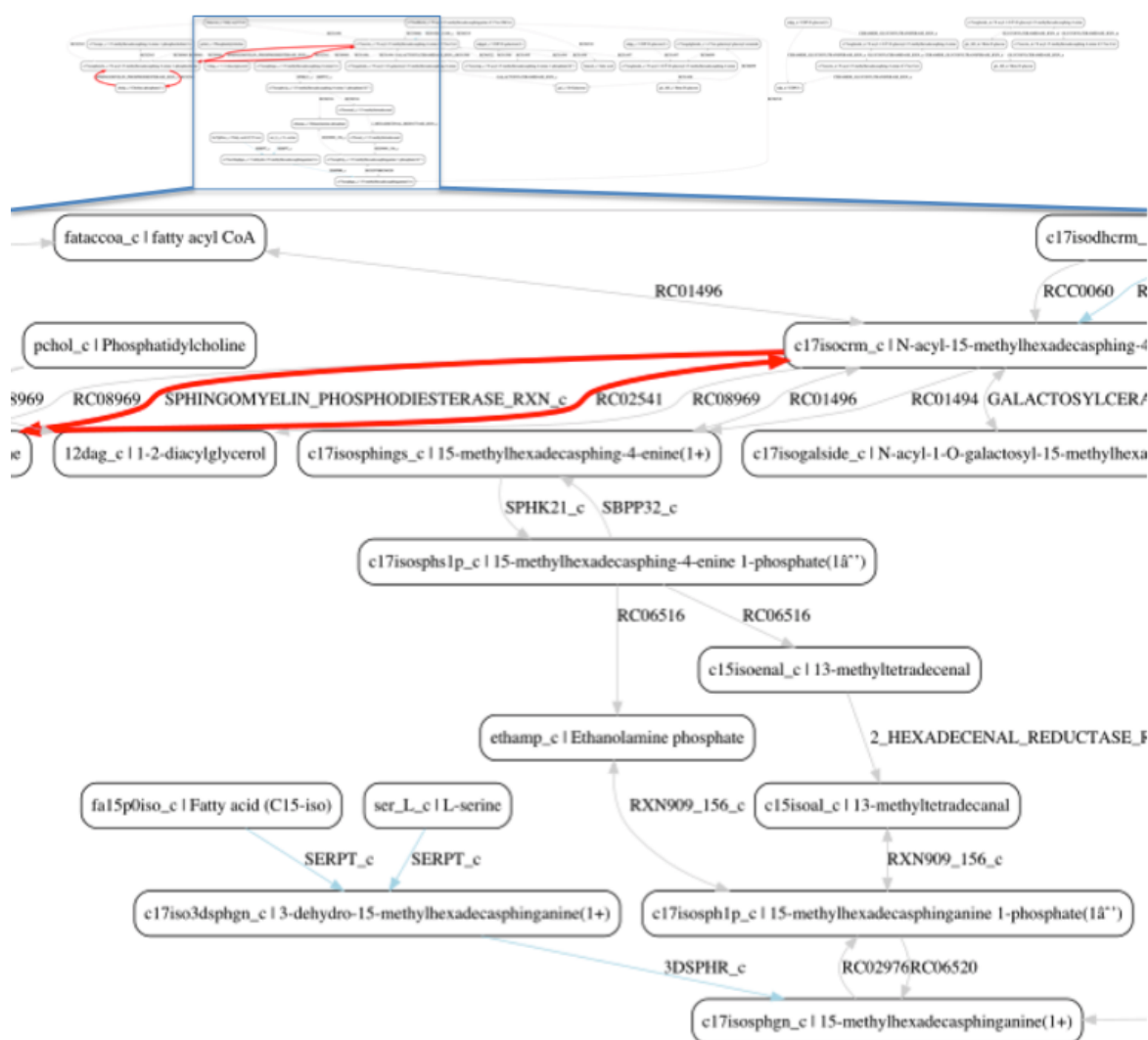


Figure 3.6 Visualisation of arbitrary groups of reaction chains.

This Figure shows a set of reactions involved in sphingolipid metabolism, an aspect of the WormJam model that was recently corrected by manual curation. Each metabolite that features in the reaction chain is illustrated as a node with both ID and name displayed, while reactions are illustrated as edges connecting nodes where reactions consume one metabolite and produce another. Edge colour indicates flux magnitude, with red = large flux and blue = small flux. Zero flux is indicated as grey.

reason, I implemented my own reaction set visualisation utility, as described in the Methods. For example, a subset of reactions that was added to better describe sphingolipids in the latest release, as described in (Witting et al., 2018), is shown in Figure 3.6. While clearly this is not the most beautiful visualisation utility ever developed, the important attributes which aid in debugging of broken reaction chains include the display of both identifiers and names for the metabolites as well as reaction IDs for the interlinking reactions, as well as overlaying fluxes from a given solution onto the display of reactions with colour and arrow width. After the initial build of this version of the model, the end-product in this

set of reactions (C17-iso-sphinganine) was not able to be made by the model. It now is makeable (note the flux, i.e. a blue line for reaction 3DSPHR_c). However, it should be apparent from the large number of grey lines representing reactions with no flux that this part of the model is still not optimally healthy and will require further curation to resolve all blocked intermediary components.

3.3.4 The WormJam consensus reconstruction: current status

As mentioned above, the most recent release version of the WormJam model is available online at the WormJam GitHub repository. The size of the model is given by counts of various component parts in Table 3.2, and an overview visualisation of the model generated with the Cytoscape network visualisation tool (Su et al., 2014) is shown in Figure 3.7.

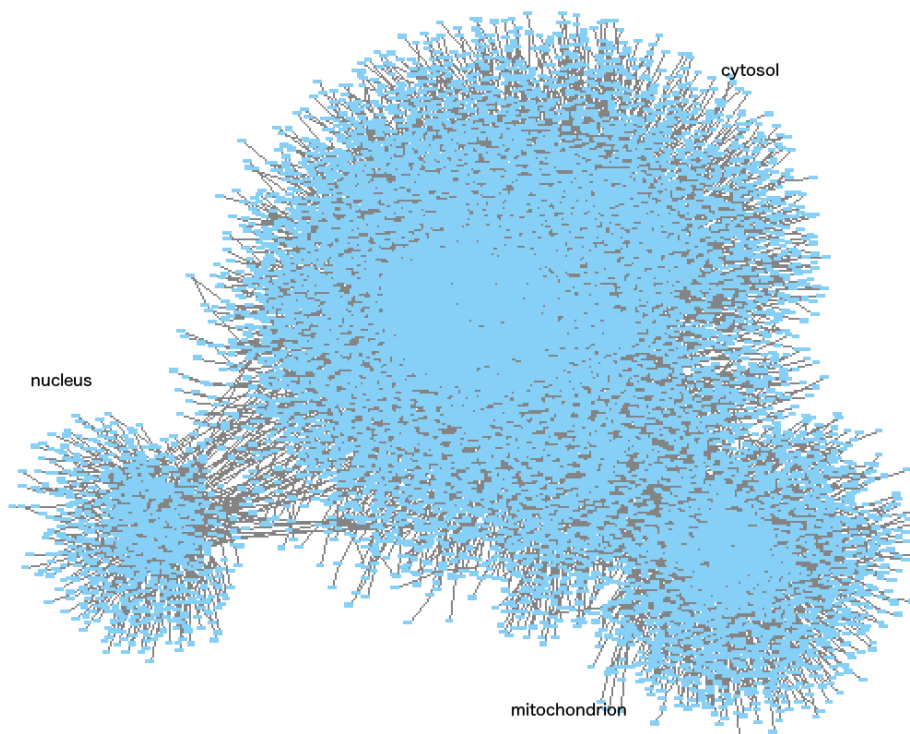


Figure 3.7 Overview of the entire WormJam consensus model.

The interconnected whole network of reactions and metabolites visualised with the Cytoscape tool. Nodes represent reactions and metabolites, edges participation of metabolites in reactions. The layout was computed automatically, but clear separate clusters for the different compartments are visible.

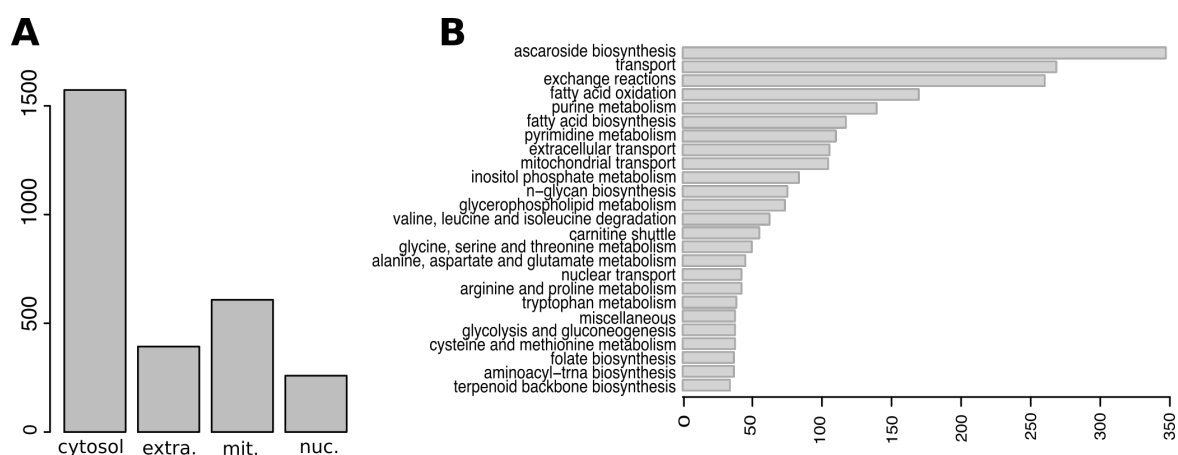
The WormJam model has more content than either of the pre-existing models (for which the metrics are given in Table 3.1). The plurality of possible food types and stages of growth that were implemented in the original models have been retained as options in the merged model. Furthermore, some of this additional content is unique to *C. elegans* and unique to

Table 3.2 Metrics for the size of the WormJam consensus reconstruction

The counts of the various constituents of the WormJam consensus reconstruction model are shown. The number of unique (across compartments) compounds is shown in brackets after the total number, as in these types of models, compounds are distinguishable per compartment. Abbreviations: cyt.=cytosol, mit.=mitochondrion, nuc.=nucleus, extra.=extraorganism/extracellular.

| Component | Size (Count) |
|--------------|------------------------------|
| Metabolites | 2834 (1630 unique) |
| Reactions | 3634 |
| Genes | 1520 |
| Pathways | 154 |
| Compartments | 4 (cyt., mit., nuc., extra.) |

the WormJam model, such as the ascaroside biosynthesis pathways and aspects of the fatty acids pathways. Thus, WormJam represents an advance on the pre-existing models for this organism.

**Figure 3.8 Sizes of compartments and pathways in WormJam model.**

(A) The count of metabolites per compartment. (B) The counts of reactions in the top 25 most populous pathways in the model. extra.=extra-organism (extracellular); mit.=mitochondrial; nuc.=nucleus.

Figure 3.8 shows some indicators for the distribution of content in the different compartments and pathways within the model. The majority of the model content is localized to the cytosol (Figure 3.8A), as would be expected firstly as this is the largest compartment in any genome-scale model of cellular metabolism, but particularly so given that many other cellular compartments are not explicitly represented (peroxisome, lysosome etc.). The counts of reactions for the top 25 most populous pathways are illustrated in Figure 3.8B, while the full set of counts for all the pathways are included as Supplementary Table A.4 in the Appendix.

The largest pathway in terms of reaction counts is ascaroside biosynthesis, a pathway that was added entirely in the last round of manual curation of the model. Various different

groups of transport reactions also form a sizeable portion of the model's content: these reactions are mostly not annotated with active enzymatic transporters, but are needed to ensure that metabolites can move around the different parts of the model as needed. Fatty acids feature prominently among the largest pathways in the model, as both fatty acid oxidation and fatty acid biosynthesis are among the top 10 pathways. This is despite the fact that fatty acids are *not* well represented in the model relative to other, simpler metabolites (Witting et al., 2018), and a more complete representation would involve a substantial expansion of the annotation of these pathways in the model.

Lipids represent a major bottleneck in metabolic models in general, not only in *C. elegans*. Lipids cover a large combinatorial space with different fatty acid combinations possible, which is hard to represent accurately in a stoichiometric model. In addition, *C. elegans* harbors several peculiarities in its lipid metabolism (Witting and Schmitt-Kopplin, 2016).

The other large pathways include purine and pyrimidine metabolism as well as metabolism related to various amino acids.

The comparison of pathway sizes can only be considered as a guide to the proportion of the underlying biology described in the model, as pathway sizes are also indicative to some extent of granularity differences in the choice of reactions to include in discrete pathways, which can be selected in a number of different ways. Indeed, one of the major differences between the two original models was in the selection and description of pathways and how these were associated with reactions. Individual reactions can also belong to multiple pathways, and some reactions are inherently bridging reactions between different pathways e.g. lipids and energy metabolism. However, for meaningful biological interpretation it is useful to minimise overlaps as far as possible, and the pathway annotations will be used extensively for the purpose of interpreting predicted fluxes in subsequent chapters.

3.3.5 Model validation

The basic validation of the model, as mentioned above, is whether it can 'grow', that is, carry flux through to biomass creation in an FBA simulation, performed as described in detail in the next chapter. This is true of the latest version of the model. However, we can go further and explore the relationship between growth and input under different input scenarios, as well as which reactions are still blocked and how well the model is able to predictively match known metabolic phenotypes. These validations are discussed in the remainder of this section.

Growth as a function of varying input

Bacterial uptake or digestion is the primary food source that is made available to the model, representing the default *C. elegans* experimental setup. As might be expected, growth of the model varies as a function of the bacterial input linearly up to a limit, where other bottlenecks start to apply. This is illustrated in Figure 3.9A. The growth rate, i.e. the amount of biomass that can be constructed per unit of time, is linearly dependent on the rate of allowed bacterial digestion, as long as all other variables are controlled, after an initial minimal threshold of input has been satisfied, hence the non-linear jump from zero to the initial starting growth rate. The actual numbers here are less relevant than the relationship between the two variables, which indicates an importantly constraining feature of the model, all else being equal.

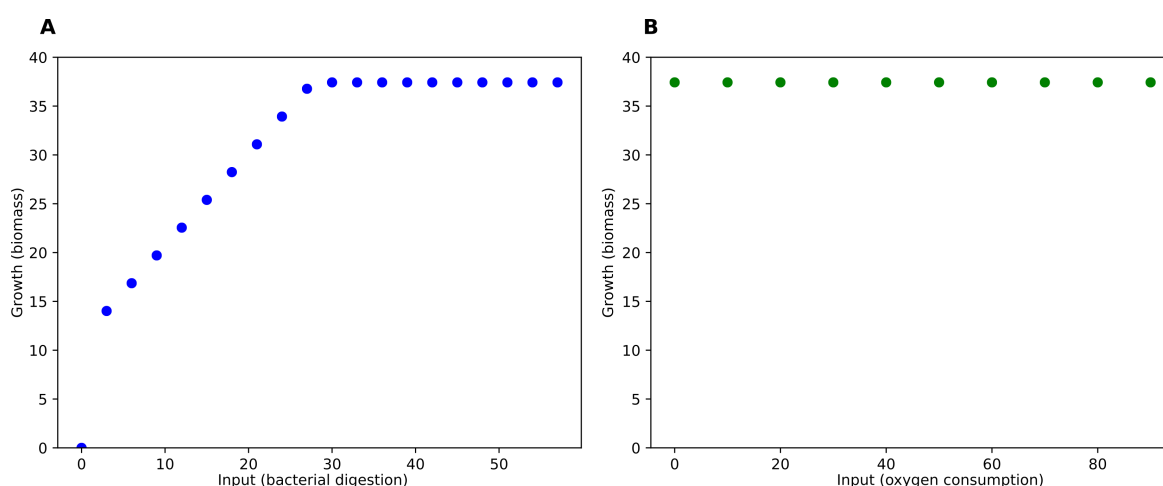


Figure 3.9 Growth of the model as a function of allowed inputs.

The growth of the model is illustrated as a function of allowed inputs (A) bacterial digestion, and (B) oxygen consumption. Growth is predicted by FBA, using a method that will be described in full in the subsequent chapter.

The rate of growth of the model is, however, *not* dependent on the allowed rate of oxygen consumption (Figure 3.9B is absolutely flat). This is a bit worrying, as ordinarily models of this type would be expected to grow more under aerobic conditions than anaerobic, however, in this model under default conditions it seems to have no effect.

On the one hand, this result can be interpreted to reflect only that *C. elegans* is able to harness anaerobic metabolism, which is true. In fact, *C. elegans* has even been shown to be long-lived under certain hypoxic conditions (Ma et al., 2017), reflecting that this is to some extent a biochemically viable result.

On the other hand, the result may also reflect that the model is under-constrained and too permissive with respect to alternative pathways for energy generation. However, we do see that the rate of oxidative phosphorylation (ATP synthase) is connected to the rate

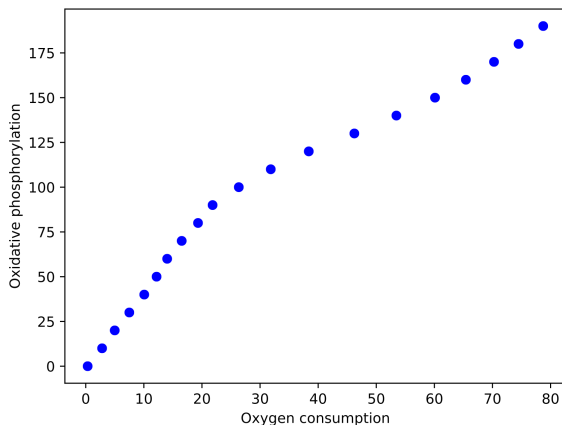


Figure 3.10 The rate of ATP synthesis from ADP via oxidative phosphorylation in the mitochondrial compartment in the model is illustrated for different levels of oxygen consumption. ATP synthesis is reaction 'ATPS4m' and oxygen consumption is reaction 'O2_exchange_reactions_e' in the WormJam model. For this study, parsimonious FBA was used which tries to minimize the overall flux sum.

of oxygen consumption (Figure 3.10), but only if a 'parsimonious' FBA implementation is used – another method that will be presented in full in the next chapter – which tries to minimize the overall sum of fluxes in the solution. Overall, it is likely that further optimization of the model's energy generating pathways will be needed in order to better represent physiological realities.

Blocked reactions

As mentioned above, in large-scale metabolic models of this type, it is normal for a certain percentage of the reactions in the model to be 'blocked', that is, not able to carry flux in an FBA simulation, under typical conditions. Some of the blocked reactions would be able to carry flux in alternate conditions, such as alternative input nutrients (food) or alternative biomass compositions. However, the present evaluation relates only to a standard scenario of inputs and outputs.

Figure 3.11 shows the percentages of blocked reactions in particular pathways, ranked from the highest (100% blocked reactions in a pathway) to the lowest (less than 10% blocked reactions). For ease of visualisation, the Figure includes only those pathways that include more than ten reactions, however, the full list of reactions together with the counts of blocked reactions is listed in Supplementary Table A.4. There are of course also pathways with no reactions blocked at all, which are not illustrated in the figure but are included in the supplementary table.

Several pathways are fully blocked, that is, they are never being used under ordinary conditions when the model is used for simulations. The fully blocked pathways include straight-chain fatty acid synthesis, steroid and steroid hormone biosynthesis, primary bile acid biosynthesis, phosphatidylinositol phosphate metabolism, core modifications of n-glycans, and the carnitine shuttle. It is clear that major aspects of fatty acid and secondary metabolite biosynthesis are not fully functional in the model, a shortcoming that is shared

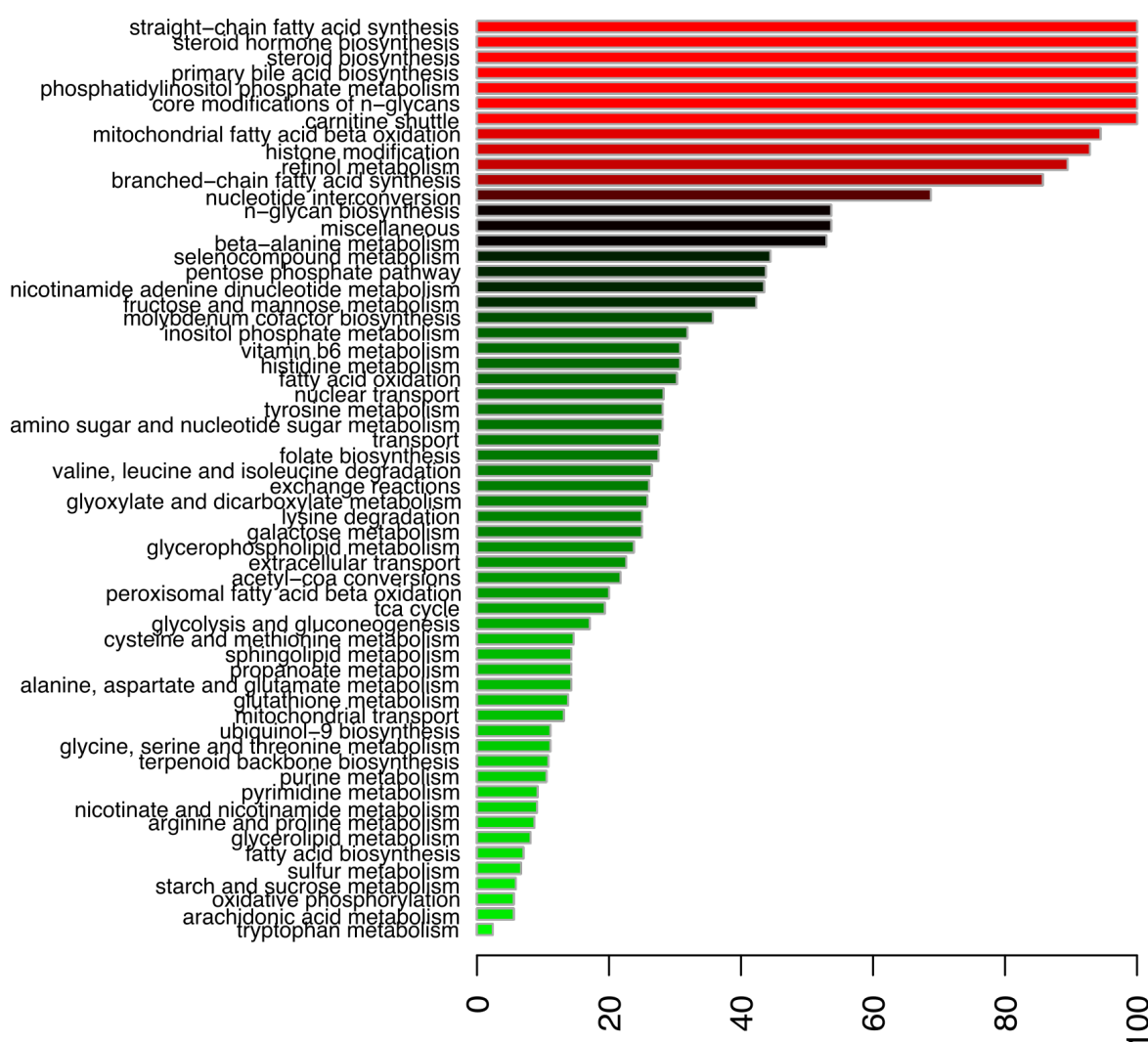


Figure 3.11 Percentages of blocked reactions in model pathways.

The percentage of reactions that are blocked in the different pathways of the model are illustrated in descending order, with the highest percentages coloured in red while the lowest percentages are coloured in green. This Figure includes only pathways that have more than 10 reactions in total and at least one blocked reaction. The full list of pathways is included in Supplementary Table A.4.

with many other genome-scale models, as also discussed in (Witting et al., 2018). Some aspects of fatty acid metabolism are better represented than others. For example, while mitochondrial fatty acid beta oxidation is close to fully blocked, interestingly, peroxisomal fatty acid beta oxidation is in much better health, at around 20% blocked reactions. Fatty acid biosynthesis is also well in the green, as is sphingolipid metabolism, a particular focus of the manual curation and debugging phases of the last round of model curation. Core central metabolic pathways such as the TCA cycle, glycolysis and gluconeogenesis, purine and pyrimidine metabolism are all well under 20%, reflecting the better state of annotation of core metabolic pathways.

Comparison of essential genes to known lethal genes

One standard approach to validate models of this type is to compare the genes that have a direct effect on the growth of the model with genes that are known to have lethal phenotypes when knocked out. The idea is that the genes that are essential in the model should correspond as far as possible to the genes that are known to be lethal, since growth of the model corresponds to functional metabolism. Both of the originally published models ElegCyc and iCEL1273 validated their model's ability to predict genes with known lethal phenotypes, although they used different sets of genes and different approaches in their validations.

Taking a simple approach and just comparing the list of essential genes predicted by the WormJam model (included as Supplementary Table A.5 in the Appendix) for standard inputs and biomass, with known lethal gene deletion mutants from the WormBase RNAi database, we found that only 18 of the model-predicted essential genes were known lethal in the RNAi dataset. There were 105 genes predicted as essential by the model that were not listed as having lethal phenotypes based on RNAi experiments (false positives). On the other side, 36 genes that are known to have lethal phenotypes were not predicted as essential for the model (false negatives). Among the false positives are predicted homologs of human tRNA synthetases, e.g. *lars-2* (leucyl tRNA synthetase) and *gars-1* (glycine tRNA ligase), which are required for the model representation of the composition of the RNA component of biomass. Among the false negatives are *pcyt-1*, a choline-phosphate cytidyltransferase involved in the synthesis of phosphatidylcholine. This gene is associated with just one reaction in the model, choline-phosphate cytidyltransferase as would be expected. However, that reaction is annotated with two genes associated with OR logic, that is, either of the genes can catalyse the reaction, thus knocking out either of the genes in a single gene deletion test would not knock out the reaction. The other annotation, F28A10.10, is listed in WormBase as a non-coding pseudogene, meaning that the annotation is probably legacy and should be removed. However, in this case, even if the reaction itself is knocked out, the model is still able to grow, indicating that choline-phosphate cytidyltransferase activity is not essential for the metabolism as represented in the current biomass requirement, since phosphatidylcholine can be acquired from the bacterial food source.

There are many ways in which such a simple analysis can be improved. Firstly, the WormBase RNAi dataset is not a perfect source of lethal phenotype predictions as the experiments may have been done at different times and in different genetic backgrounds. To mitigate this challenge, an experimental collaborator has prepared a screen testing all of the WormJam genes under precisely the same conditions and with the same genetic background, for which we are awaiting the results. Secondly, the effects of such an intervention may be direct or indirect, affecting other genes through regulatory mechanisms, which will not

be accessible to the metabolic model. Finally, many of the gene-product associations to reactions are conservative with many possible genes listed with OR annotations, and in these cases knocking out any one of the genes will have no effect on model output, thus, a different algorithm which marked a reaction as knocked out if any of the associated genes are unavailable would have possibly better results – more false negatives, but possibly more true positives.

The issue with conservative annotation in the absence of definitive biochemical knowledge extends also to redundancies created by flagging reactions as reversible. Just over a third (1,235 reactions, which is 34.9% of the total) of the reactions in the WormJam model are currently annotated as reversible. Some of these, such as transport reactions representing the movement of metabolites across membranes, indeed are fully reversible. Many others are biochemically reversible, but the reverse reaction is catalysed by a different enzyme, a fact that is not fully captured by the ‘reversibility’ model shorthand. Other annotations are simply wrong in that the reactions may be reversible *in vitro*, but the reverse reaction would never carry a significant flux *in vivo* as it may be energetically infeasible, for example. Indeed, fundamental metabolic regulatory controls and segregation of cellular locations have evolved precisely in order to avoid spurious reverse metabolic reactions which would be wasteful. Thus, improving this aspect of the annotation of the model should also improve the correspondence of the model predictions to physiological phenotypes.

Comparison of metabolomics studies to metabolites in model

In (Witting et al., 2018) we reported a comparison of the metabolites that are annotated in the model with the metabolites that have been detected in previously published *C. elegans* metabolomics studies. Several groups of metabolites were found to be missing in the model, highlighting areas where future curation will be important to ‘plug’ the gaps.

Fatty acids were a group that was found to be the most under-represented in the model, in common with other genome-scale models. Fatty acids are the building blocks of lipids, and are also important for the synthesis of secondary messengers and signalling molecules. Very long chain fatty acids with chain lengths up to 30:0 have recently been detected in *C. elegans* (Gao et al., 2017, 2018). However, even if detailed biosynthesis pathways for each individual lipid could be integrated into the model, there would still be a discrepancy between detailed structure in the model and the annotation and identification capabilities of current lipid analysis methods. With standard lipidomics analysis, usually neither the position nor the stereochemistry of double bonds can accurately be identified. Therefore, new approaches are needed to make it possible in the future to incorporate lipidomics data.

Relatedly, several new ascarosides were identified in the literature, containing very long fatty acid side chains, which are currently not covered in the consensus model. There is

growing evidence that the biosynthesis of ascarosides is more complex than previously expected (Zhou et al., 2018). So far, only simple ascarosides have been annotated in the model, but ascaroside biosynthesis already makes up the largest pathway with respect to the number of reactions it includes (Figure 3.8B).

Several modified nucleotides have been detected in *C. elegans* (van Delft et al., 2017), including 1-methylguanosine, 1,7-dimethylguanosine, 1-methylinosine, 5-methylcytidine, and *N*6-carbamoyl-L-threonyl-adenosine, representing modified RNA degradation products. These modifications play important roles in fine tuning RNA function, and their production consumes metabolites such as SAM, thus representing an important sink for methyl groups possibly competing with other cellular metabolic and epigenetic functions. They will need to be added in future versions of the model.

We also compared the metabolites in the model to those included in our metabolomics dataset as a part of the mapping to use the model together with our dataset, and this is described in full in the next chapter.

3.4 Conclusion

While the current WormJam reconstruction is a good consensus between existing efforts, and the most comprehensive *C. elegans* metabolic model to date, the task of reconstructing the metabolism of the worm is by no means finished. The accuracy of the reconstruction is not uniform. For example, the metabolism of carbohydrates and peptides is fairly well described, but the coverage of nucleic acids and lipids needs improvement. The number of different compartments is limited, and in particular some organelles featuring very specific biochemistry, such as lysosomes and peroxisomes, are not explicitly represented. Moreover, *C. elegans* is a multi-cellular organism interacting with a complex environment, and its cellular metabolism cannot be understood in isolation. Interactions between tissues and with the active metabolism of the bacterial food source will need to be annotated in the future.

Community-based curation and refinement of the WormJam model promises to address several of the challenges facing the field. It is essential that annotation of metabolites, reactions, and of gene associations to reactions, be improved – and to record the confidence associated with such annotations to enable more sophisticated algorithms to be developed that are able to prioritise ‘core’, well annotated parts of the model in evaluating predictions.

With the curation undertaken thus far, the first steps have been taken toward an improved representation of *C. elegans*-specific metabolism and biology. However, the network still contains redundancy, and major parts are still missing, e.g., better biomass

compositions for constraint-based modelling including detailed fatty acid profiles for individual lipid classes, known to be highly important for ageing and longevity.

However, despite all the areas which still need improvement, the model is already usable and may provide useful predictions despite its limitations. In the next chapter, we put the model to work predicting reaction fluxes corresponding to our time series of -omics measurements and determine whether it can offer any useful insights into *C. elegans* ageing.

4. Metabolic flux changes with ageing

4.1 Introduction

Metabolic fluxes are the rates of turnover of molecules through specific reactions, modulo the reaction stoichiometry. They represent a quantity of biochemical reaction events per unit of time, that when multiplied by the stoichiometries for each molecule that is consumed or produced, provide the rates of molecule consumption or production.

Metabolic fluxes are tightly controlled via direct and indirect regulatory mechanisms, and are responsive to different environmental conditions. The fluxes through central carbon metabolism are the most tightly regulated to balance cellular energy and maintenance needs. Reaction fluxes are the outcome of gene expression, translation, post-translational modifications and protein-metabolite interactions, and can be considered to be one of the closest readouts of cellular phenotypes.

As previously mentioned, Flux Balance Analysis (FBA) is a metabolic modelling approach that uses a whole-genome model of metabolism together with various assumptions and constraints in order to predict intracellular fluxes through every reaction in the model. Because FBA calculates the flow of metabolites throughout the global metabolic network, it allows the elucidation of how changes in one aspect of metabolism affect other pathways and phenotypes (such as overall growth rate, or the rate of production of a certain metabolite), and in this way it can provide valuable mechanistic insights (Orth et al., 2010).

In this chapter I introduce a novel FBA multi-omics integrative method that I developed to predict fluxes using a linked time series of both metabolomics and the transcriptomics datasets. This method was published in (Hastings et al., 2019a), and some of the figures presented in the relevant section below are taken from that publication (acknowledged where appropriate in the Figure legend). I then go on to use this novel method to evaluate and discuss the detailed dynamic changes in metabolic fluxes that are occurring over time and as compared between the different strains in my dataset, findings which were not incorporated into the publication.

4.2 Methods

4.2.1 Flux Balance Analysis

Flux balance analysis (FBA; Orth et al., 2010; illustrated in Figure 4.1) is a mathematical modelling technique that uses reaction stoichiometry from a whole-genome metabolic

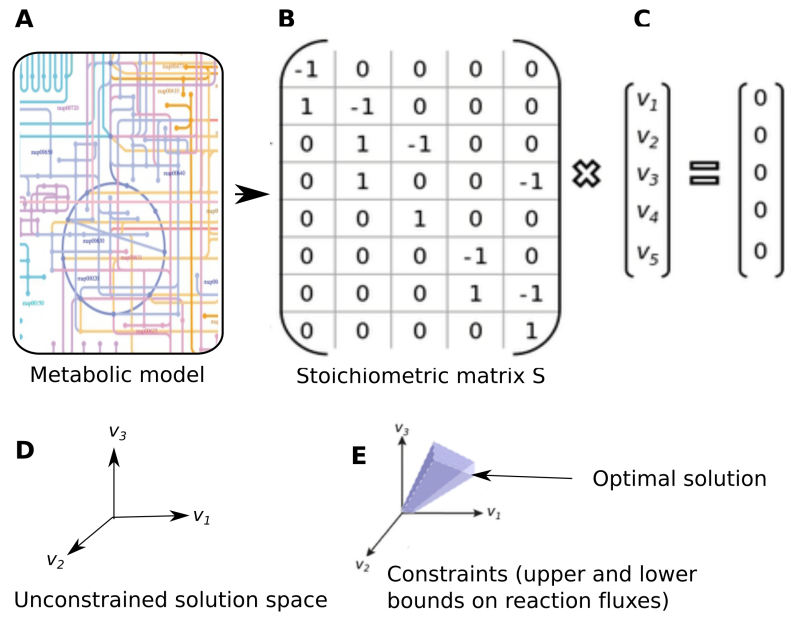


Figure 4.1 A schematic illustrating Flux Balance Analysis.

A whole-genome metabolic model (A) is transformed into a stoichiometric matrix S (B) with rows for metabolites M and columns for reactions R . The product of the matrix S with the flux vector v is 0 (steady state assumption). The solution space for this problem is still infinite (D), but together with additional reaction constraints (E), it is possible to select an optimal solution.

model, together with linear optimization and an assumption of steady state metabolism, to generate predictions of fluxes through each reaction in the full metabolic network for a specific organism.

The basic mathematical problem is formulated as follows. Given a set of metabolites $m_i \in M$ which are involved in a set of reactions $R_i \in R$ with stoichiometry matrix S , and assuming that the system is at a steady state, i.e. all metabolites which are produced are consumed at the same rate, we can say that the product of the vector of fluxes v through the reactions and the stoichiometric matrix S is 0: $S \cdot v = 0$.

This problem clearly still has a vast and underconstrained solution space, since the only information included thus far are the reaction stoichiometries and the fact that the system is at steady state. To include knowledge about the physiological constraints on various reactions (where known), upper and lower bounds may be set on the allowable flux for specific reactions. That is, for each v_i in v (the set of reaction fluxes), the model specifies a lower bound L_i and an upper bound U_i such that $L_i < v_i < U_i$.

Although these constraints reduce the size of the solution space, given these constraints alone, the solution space for any given metabolic model is still infinite. To arrive at a defined solution, it is thus also necessary to specify an objective for the solver to aim for maximising or minimising. Typically, the objective that is used in metabolic modelling is

to maximise flux v_{BIO} through the biomass reaction R_{BIO} , i.e. to maximise growth (μ). A wide range of alternative objective functions can also be used, for example to maximise the yield of various exported metabolites (a common scenario in synthetic biology) or cofactors; any reaction or combination of reactions may be optimised.

The resulting mathematical problem is then solved using linear programming. A solution of an FBA problem is a vector of quantitative flux values $v_i \in V$, with one v_i for every reaction R_i in the network. Fluxes are specified in units of mmol per gram of dry cell weight per hour ($\text{mmol.gDCW}^{-1}.\text{h}^{-1}$). This type of model includes no information about the concentrations of metabolites, only the rates at which metabolites are converted in reaction fluxes.

In what follows, I used the CobraPy library in Python (Ebrahim et al., 2013) to run FBA and as a basis for the implementation of FBA-associated methods, including integration of transcriptomics and metabolomics data (described below).

Model version and input constraints

The version of the WormJam model used for most of the analyses in the chapter was a development build dated 2018-10-12 (available from WormJam repository, as described in Chapter 3). Using a development version allowed me to incorporate the most up-to-date changes in model content and to fix errors discovered while running these simulations. However, it means that the version I used may not have been fully validated for official release. In our publication (Hastings et al., 2019a), which includes material partially overlapping with some of the analyses in this chapter, an earlier (January 2018), officially released version of the model was used. Most of the analyses reported in this chapter were re-calculated with the newer version of the model; where this is not the case it is specifically indicated.

All sources of input (uptake) were constrained to zero except:

- Oxygen (maximum of 100 units allowed uptake);
- Water (maximum of 100 units allowed uptake);
- Bacterial input as food source via bacterial digestion (maximum of 30 units allowed uptake). Bacterial input is broken down into component pools of DNA, RNA, peptides, protein, glycogen, lipids, and a pool of soluble metabolic components including cholesterol.
- Essential amino acids (Braeckman et al., 2009) (arginine, histidine, lysine, tryptophan, phenylalanine, methionine, threonine, leucine, isoleucine and valine; 5 units each of allowed uptake);

- Heme, which *C. elegans* requires exogenously (Rao et al., 2005), maximum of 10 units allowed uptake;
- Pantothenate, which is required for Coenzyme A synthesis (Lu and Balachandar, 2005), maximum of 10 units allowed uptake; and
- Trace minerals and ions (Ca^{2+} , Cu^{2+} , K^+ , Mg^{2+} , Mn^{2+} , Na^+ , Zn^{2+} ; one unit each of allowed uptake)

To some extent, these values are arbitrary, as there are few studies evaluating exactly how much of these nutrients are taken up by *C. elegans*. The limits that were chosen mostly do not constrain the system (the actual uptake is lower than the allowed uptake) and thus are chosen only to prevent the model being able to bypass ordinary energy generation pathways by overusing pathways that would not normally be physiologically possible.

The biomass reaction used as the default objective function included glycans, phospholipids, collagens, DNA, RNA, free fatty acids, glycogens, proteins, triacylglycerols, and trehalose.¹ Again, to some extent the biomass reaction is arbitrary – biomass composition has been flagged as an important limitation across genome-scale models, not just *C. elegans* (Dikicioglu et al., 2015) – and thus is an area open to improvement in the future. However, what is included is reasonably comprehensive with regard to the different molecular constituents to ensure balanced flux predictions.

FBA implementation

There are many variations and extensions on the core method of FBA, improving the underlying assumptions to move closer to the underlying biological reality, and supporting answering a wide variety of questions (Lewis et al., 2012).

Given that there are typically multiple flux states that are compatible with a given objective function, the principle of overall flux minimization, or *parsimony* is believed to give the most realistic flux values within the solution space (Holzhütter, 2004). The implementation of parsimonious FBA involves a two-step solution, where in the first instance the maximum value of the objective function is determined using linear programming as for standard FBA, and then in a subsequent second step the total flux through the matrix of reactions is minimized at the given fixed value of the objective function. An alternative strategy for parsimonious FBA is to minimise the count of active reactions rather than the total sum of fluxes (Schuetz et al., 2007). Minimising the number of active fluxes can be seen as a proxy for minimising the burden of protein expression, while minimising the sum of overall fluxes can be seen as a proxy for maximising efficiency.

¹The precise biomass reaction with stoichiometries for these components is 'BIO0100' in the model: 0.03 trehalose + 0.12 triacylglycerols + 0.06 glycans + 0.13 phospholipids + 0.19 DNA and RNA + 0.21 collagens + 0.23 glycogens + 1.15 mitochondrial protein + 4.03 other proteins + 1e-05 minor components of biomass.

Practically, I have implemented my FBA pipeline using COBRApy's parsimonious FBA method "cobra.flux_analysis.pfba".

4.2.2 Integration of -omics data with FBA

As noted above, the FBA problem is typically underconstrained, and techniques are needed to further reduce the solution space to match what is biologically feasible. Biomolecular -omics measurements can improve the predictions of FBA by constraining the solution space and by clarifying the objective for optimisation.

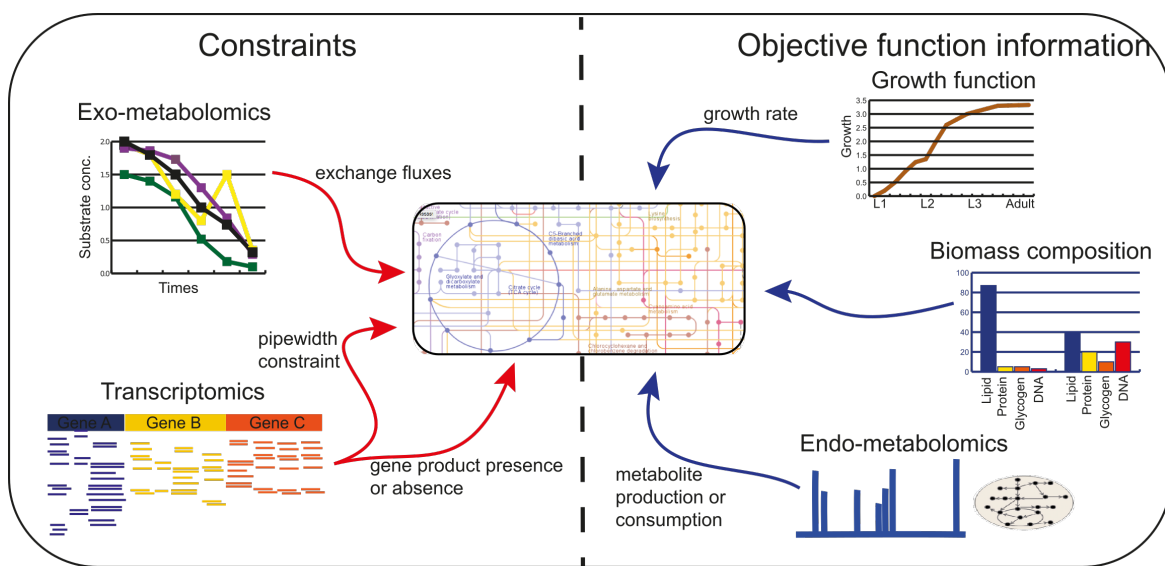


Figure 4.2 Integration of -omics data with Flux Balance Analysis.

This schematic illustration shows different ways that -omics data can be incorporated together with a whole-genome model for FBA. Methods are divided based on the data source they incorporate and whether the data is used to apply constraints to the model or to inform the objective function that will be optimised. *This Figure has been adapted with minor changes from the one presented in our publication (Hastings et al., 2019b), and was originally created by Manusnan Suriyalaksh.*

Figure 4.2 illustrates a number of different ways in which -omics data can be used together with whole-genome models to obtain more biologically relevant FBA solutions. Gene expression data and metabolomics data can both be used in different ways to constrain analyses on the network model, reducing the solution space accordingly to those that are biologically consistent with the observed data.

Each of these methods for integrating gene expression and metabolite concentration data with metabolic network models, however, relies on a mapping from the measured genes and metabolites onto the network model. Measured genes and metabolites that are not represented in the network model cannot be used by these methods. (The mapping of our measured metabolites to the WormJam metabolic model is listed in Supplementary Table A.6 in the Appendix.)

Measurements corresponding to enzyme levels, i.e. proteomics or transcriptomics, are the most straightforward to use for adding biological constraints to the FBA problem, although even this is not completely straightforward as enzyme levels and activity rates may differ, e.g. due to substrate availability etc.

Proteomics provides direct estimates of enzyme concentrations, although not necessarily of the active form, and not for all enzymes. Transcriptomics data tends to be more comprehensive, but transcriptional readouts are only a proxy for enzyme concentration.

There are multiple tools to use gene expression data together with metabolic network models (Lewis et al., 2012), and a systematic evaluation has found that none of the methods outperforms all others for all conditions, but rather that different methods show strengths and weaknesses in different scenarios (Machado and Herrgård, 2014), while keeping in mind that the overall quantitative correlation between gene expression levels and metabolic flux catalysed by the encoded gene products is known to be low (Yang et al., 2002).

In the earliest such methods developed (Akesson et al., 2004; Becker and Palsson, 2008), the problem of the low correlation between gene expression and metabolic flux was avoided by having the integration hinge on whether or not a particular gene was expressed at all (i.e. gene expression was binarised). If a particular gene was not expressed (Akesson et al., 2004), or was expressed below a given threshold (Becker and Palsson, 2008), then flux through any reactions which depend on that enzyme were constrained to zero (shown as 'gene product presence or absence' in Figure 4.2). This results in a reduced size, context-specific metabolic network consistent with the genes known to be expressed in the given condition.

Moving towards a more quantitative use of gene expression data, Colijn et al. (2009) introduce the E-Flux method where the presence of higher enzyme levels is taken to imply that more of the reaction (flux) can take place, corresponding to a 'pipe size'. In this integration approach, maximum constraints for enzyme-catalysed reactions in the model are set in proportion to measured levels of enzyme expression (shown as 'pipewidth constraint' in Figure 4.2).

Using even more of the quantitative information available in the gene expression, the iMAT method (Shlomi et al., 2008; Zur et al., 2010) tries to make the obtained flux solution proportional to the measured enzyme levels by using both low gene expression and high gene expression in a given condition to constrain the solution of the model. Fluxes are minimized through parts of the network that are lowly expressed, and maximized through parts of the network that are highly expressed.

Other methods exist which use alternative approaches to incorporate quantitative information from gene expression to constrain fluxes in the network, including fold changes between differential conditions (Jensen and Papin, 2011).

In the pipeline developed for this study, the available transcriptomics data were first log-transformed and then means were obtained for biological replicates in order to have a single expression value per time point and strain. These were integrated with the model using a two-step approach. First, a context-specific model was built by removing reactions mediated by genes whose expression was below a threshold (gene product presence/absence, Akesson et al., 2004). The threshold used was arbitrarily set to 0.8 in log-transformed read counts. Then, for the remaining reactions, the E-flux approach (Colijn et al., 2009) was used, in which reaction upper (and lower if the reaction is reversible) bounds are set in proportion to the expression level. Where multiple genes were annotated to a reaction using OR logic (i.e. any of the annotated genes may catalyse the reaction), their expression levels were added together to determine the reaction bound. For complexes, where multiple genes are annotated to the reaction using AND logic (i.e., all of the annotated genes are required to catalyse the reaction), the minimum of their expression levels was used.

4.2.3 Differential flux pathway analysis

Once a set of reaction fluxes have been obtained from an FBA pipeline, these values still need to be interpreted and their biological significance determined.

Which flux differences are significant?

Comparisons between predicted flux values are challenging due to the nature of the flux data. Flux values are not normally distributed, and flux values may be either positive or negative (negative values are a convention reflecting that for a reversible reaction, the reverse reaction is in fact taking place and not the forwards reaction).

Absolute flux values (and consequently absolute flux differences) are known to vary in range dramatically for different parts of the reaction network (Almaas et al., 2004), further hindering determination of which flux differences can be considered 'significant' when comparing between conditions, since it cannot be assumed that reaction fluxes that are more different are *prima facie* thereby more interesting.

To associate some form of significance to the flux predictions, I therefore predicted fluxes separately for each of the transcriptomics replicates in each study group. For each condition in our dataset there were several flux predictions available per condition, and a statistical comparison (a simple student's t-test) was then used on the resulting fluxes to determine which of the differences were significant. The datasets of replicates flux predictions were normalised by growth rate (biomass production, per sample) prior to performing comparisons, as otherwise almost all flux-carrying reactions differ since many reactions are coupled indirectly or directly to the growth rate.

The output of the flux comparison is a list of reactions that can be considered significantly different between conditions. The t-test is not a robust statistic for this dataset, since as mentioned the flux values are not normally distributed, but the underlying read counts on which the constraints for the flux predictions are based are normally distributed, and it does have heuristic value in that it predicts a reasonable subset of flux differences are significant. I tried a non-parametric alternative statistic (Wilcoxon-Mann-Whitney) and was not able to determine *any* significant flux differences with that statistic.

Moreover, I also tried flux sampling from the solution space, an approach that has previously been suggested as a method to address flux comparisons between conditions. Flux sampling as implemented in CobraPy is based on (Megchelenbrink et al., 2014). However, flux sampling is intended to sample the *full* solution space, not just the optimal solution space. As the solution space is large, the flux solutions selected by flux sampling were very broad and varied dramatically from the optimal solution selected by optimisation in conjunction with the principle of parsimony.

Thus, significance was associated to comparisons between reaction fluxes based on the predicted fluxes for the replicates. To further validate this approach, I checked that the replicates do indeed cluster together in a PCA analysis, which they largely do (Supplementary Figure A.6 in the Appendix).

Which pathways have significant differences?

Using the list of relevant differences in reaction fluxes between conditions, to investigate which pathways are most responsible for the changes observed between different conditions (strains and times), I used Fisher's exact test for category over-representation (already introduced in Chapter 2) to identify which pathways included more different fluxes between two given conditions than would be expected by chance alone, when tested against the background list of all reactions in the model.

Fold enrichment for each enriched pathway, a quantitative measure of *how* enriched the pathways is aside from the *statistical* significance of the enrichment, is calculated standardly as it is in other category enrichment analysis tools such as (Zhang et al., 2005), as the ratio of the two proportions:

$$\text{fold enrichment} = \frac{m/n}{M/N} \quad (4.1)$$

where m is the number of significantly changed reactions in the pathway of interest, n is the number of reactions that are significantly changed overall (all pathways), M is the number of reactions in a given pathway, and N is the total number of reactions in the whole model.

4.2.4 Flux visualisation

In this chapter, predicted fluxes are mainly visualised either directly, with simple bar plots, or in the context of specific reaction groups with the Escher pathway flux visualisation utility (King et al., 2015). Escher relies on hand-drawn maps of metabolic reactions, and offers options to colour reactions by a single set of flux values or by a comparison of two sets of flux values. Where a flux comparison is visualised, reaction colours indicate the magnitude of the differences in flux values rather than absolute flux values.

4.3 Results and Discussion

4.3.1 A novel method for integrating metabolomics data with FBA

Motivation: Standard FBA does not accurately recapitulate fluxes measured *in vivo*

We performed standard FBA, i.e. FBA with integrated transcriptomics data as described in the Methods. As objective function, we maximised biomass production (i.e., growth). When using differential flux pathway enrichment to detect which pathways were significantly changed in day 10 compared to day 5 FEM animals (Figure 4.3), we found that some pathways that are expected to change with age were enriched, such as fatty acid metabolism, glycolysis and gluconeogenesis, and growth rate (biomass assembly). Similar predictions were reached in previous studies where FBA was applied to ageing networks (Gebauer et al., 2016), indicating that standard FBA can broadly recapitulate overall metabolic functionality.

However, some pathways that are known to change with age are noticeably missing from the list, including the TCA cycle, which is amongst the most dramatic observed losses of function with age, yet (as discussed in Chapter 2) a significant component of these changes appear to be post-transcriptional and are therefore not necessarily reflected in the transcriptomics dataset. Looking at the differences between standard FBA-predicted fluxes for the FEM animals at day 5 and 10, (Figure 4.4), we can see that the differences are largely minor (blue arrows) except for a switch from the NAD-consuming isocitrate dehydrogenase reaction to the NADP-consuming isocitrate dehydrogenase reaction, which may be merely a stochastic switch in cofactor preference.

Thus, a standard FBA pipeline encompassing an objective function geared only towards biomass production, and constraints derived only from transcriptomics data, does not appropriately reflect the biology of post-mitotic somatic cells in adult ageing animals. To obtain more accurate flux predictions during ageing, we enhanced our FBA pipeline by integrating our linked metabolomics dataset into the objective function together with the biomass production. We christened this novel method ‘Metab FBA’.

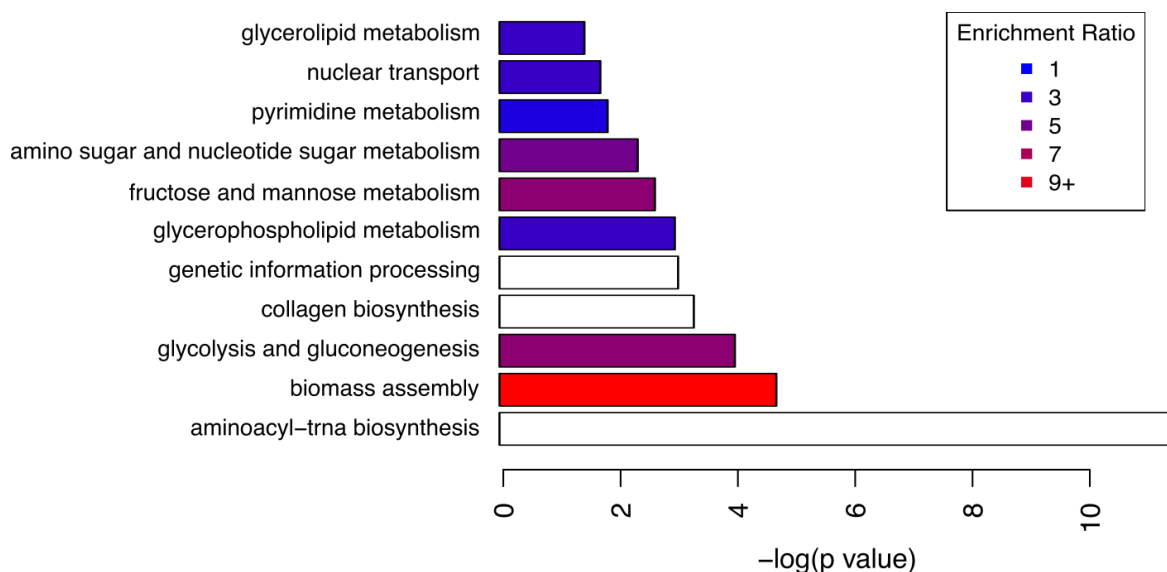


Figure 4.3 Pathway enrichment comparing fluxes between days 5 and 10 in standard FBA

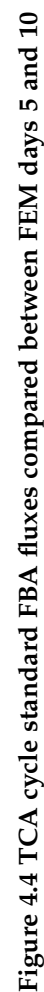
Differential pathway enrichment results are shown for the comparison between standard FBA predicted fluxes at days 5 and 10 in FEM animals. Bar size shows the negative log of the p value as an indication of the significance of the enrichment, while the category enrichment ratio, shown by bar colour, indicates the magnitude of the enrichment in terms of the proportion of affected reactions.

Method: Metab FBA

It is less straightforward to combine metabolomics data with FBA than gene expression data, as there is no straightforward correspondence between fluxes in FBA and measured metabolite concentration levels (Töpfer et al., 2014, 2018). However, several approaches have been developed which harness metabolomics data for this purpose in different ways, and the approach we have developed builds on these.

GIM3E (Schmidt et al., 2013) is an approach that relates metabolic concentration data to the model by creating ‘sink’ reactions (similar to biomass components) for measured metabolites and forcing flux through those reactions to be non-zero. Metabolomics data can also be used to inform the biomass composition or objective function (Töpfer et al., 2015).

Other approaches require a time series of measurements so that a *change* in metabolite level, rather than a direct measured value, can be related to the appropriate flux values. Where such measurements correspond to *extracellular* metabolite levels (i.e. exometabolomics), they can be directly used as constraints on the input and output (exchange reaction constraints) for dynamic FBA approaches. Exometabolomic measurements can be used to constrain the model uptake rates by determining the net consumption between two time points to give a value for nutrients per unit of dry cell weight, which is the flux unit. Typically, nutrients in the extracellular environment decrease (are consumed), and then the



The comparison of standard FBA fluxes through the TCA cycle, visualised in the Escher tool for the difference between fluxes in FEM animals at days 10 and 5. Grey arrows indicate no difference in flux, blue indicates small differences only, while pink or red indicates large differences in compared fluxes. The colours do not provide information about whether the fluxes were increased or decreased, just the absolute magnitude of their difference.

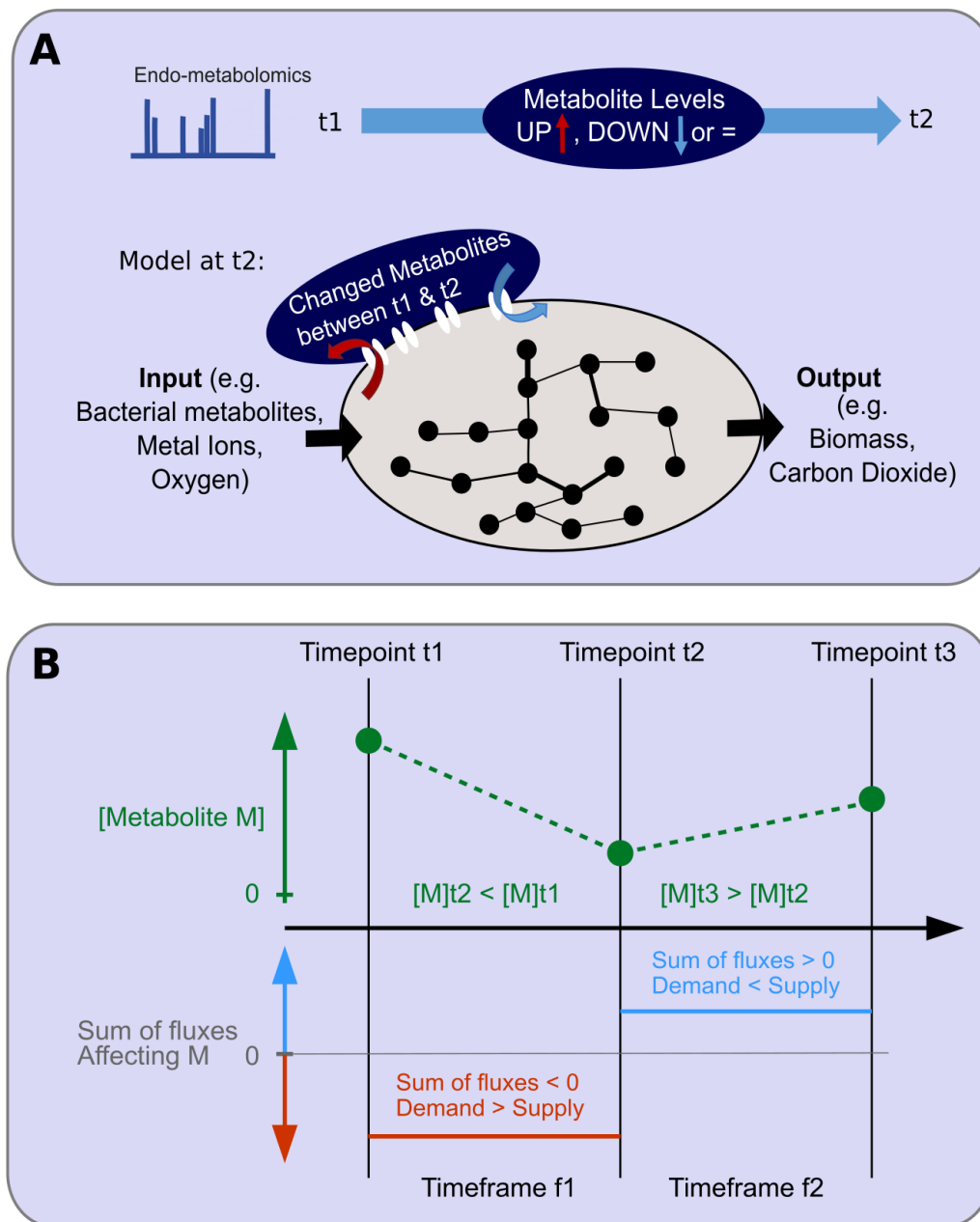


Figure 4.5 Overview of Metab FBA approach

(A) Schematic overview of the Metab FBA approach, where t1 and t2 are two consecutive time points ($t_2 > t_1$). (B) Schematic representation of the underlying assumption that metabolite level changes between time points correspond to sustained differences in fluxes within the time frame. This Figure has been incorporated from our publication (Hastings et al., 2019a), and was originally created by Bhupinder Virk and Sharlene Murdoch.

net uptake flux is negative; in case the nutrients or byproducts have seen a net excretion the uptake flux bound is positive.

In MetDFBA (Willemsen et al., 2015) and similarly in TREM-Flux (Kleessen et al., 2015), a time series of measured metabolite levels are used to determine changes in metabolite concentration levels between time points. These between-time-point changes are then used as constraints on the reaction fluxes for reactions involving those metabolites in a dynamic FBA time series. A recent approach, called ‘unsteady state’ FBA (Bordbar et al., 2017), goes beyond comparing individual time points by fitting a linear function to a series of measured values over a period of time to obtain a rate of change for a given metabolite over time, and uses this information to constrain allowed departures from steady state between time points.

The approach we introduce, Metab FBA, is similar to TREM-flux and ‘unsteady state’ FBA in that we require a time series of metabolomics measurements, but in contrast to those approaches we do not use the metabolomics values to constrain the model, but rather to alter the objective function. While ordinarily, the objective function for FBA is linked to growth, for multicellular organisms, growth can no longer be assumed to be the sole cellular objective and other phenotypically appropriate proxies are needed (Hastings et al., 2019b). We consider that, from a cellular perspective, if a metabolite level has increased between two time points, then this implies that within that timeframe the sum of fluxes through all reactions producing that metabolite (i.e. its supply) must have exceeded the sum of the fluxes through all reactions consuming it (i.e. its demand). Conversely, if the metabolite level decreased, then within that timeframe, the demand must have exceeded the supply (Figure 4.5). Accordingly, we added the net-production or net-consumption of changed metabolites as an additional objective function alongside the production of biomass, simultaneously maximizing biomass and production or consumption of the relevant metabolites.

We chose this approach because our first experiments using standard FBA on this dataset showed that using a constraints-based metabolomics data integration approach would lead to the model not being able to grow at some of our time points, as incompatible constraints are generated (Figure 4.6A). We observed that incorporation as constraints, in some cases, narrows the solution space too much, due to incompatible metabolomics and transcriptomics constraints, leading to the inability of the model to generate biomass under that setup, e.g. time point 65 hours in Figure 4.6A. This highlights that incorporation of the metabolomics information as objective function gives the maximal flexibility to the system to find the best possible solution, maximizing congruence between the transcriptomics and metabolomics datasets.

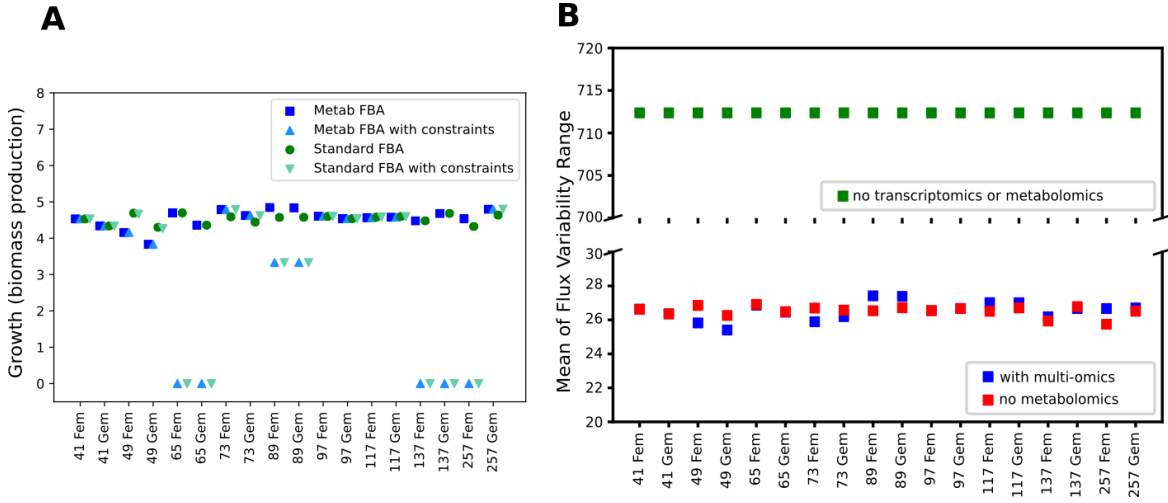


Figure 4.6 Comparison of overall indicators for Standard and Metab FBA methods

(A) Growth values for different times points when metabolomics data is used as a constraint or as objective function. (B) Mean of predicted flux variability range for the standard and Metab FBA methods. *This Figure has been incorporated from our publication (Hastings et al., 2019a).*

Our method can be formulated mathematically as follows:

$$\forall t \in (2 \leq t \leq N), \text{ maximise } (c^T v + u^T v - d^T v) \quad (4.2)$$

$$\text{s.t. } Sv = 0, \quad (4.3)$$

$$tr^L \leq v \leq tr^U, \quad (4.4)$$

$$u^T = \forall m \in M, 1 \text{ if } x^{m_t} > x^{m_{t-1}}, \text{ else } 0, \quad (4.5)$$

$$d^T = \forall m \in M, 1 \text{ if } x^{m_t} < x^{m_{t-1}}, \text{ else } 0 \quad (4.6)$$

where N is the total number of time points, t is the index of a particular time point, c^T is the vector of coefficients of the biomass equation, u^T is the vector of coefficients for demand reactions corresponding to the metabolites with level increases between time points t and $t - 1$, d^T is the vector of coefficients for demand reactions corresponding to the metabolites with level decreases between time points t and $t - 1$, S is the stoichiometric matrix for all reactions in the model, v is the vector of fluxes, M is the set of all measured metabolites that could be mapped to the model, x^{m_t} is a metabolite level for a particular metabolite m at a given time point t , and tr^L and tr^U are flux lower and upper constraints set from the transcriptomics data using the method described above.

For each metabolite per strain, a t-test was used to determine which metabolite levels had changed significantly between consecutive time points. A comparison of the means was then used to determine if the metabolites with level changes had increased or decreased. We were only able to do this for $(N - 1)$ of the N time points; flux predictions at the initial

time point are thus unaffected by the metabolomics data. Only metabolites which could be mapped to model metabolites were included; this was the case for 78 of the 105 measured metabolites. Of those, 70 metabolites had at least one significant between-time-point difference, while 8 were unchanged across all time points.

Validation: Improved prediction of age-related changes in TCA cycle

Previous approaches to incorporate -omics data with FBA have been validated by confirming that the overall variability of flux predictions – the sum of the range of flux variabilities for each reaction – is reduced (e.g. Kleessen et al. (2015)). Flux variability is a measure of how constrained a solution space is in an FBA problem, and it will be introduced in full in the next chapter where we use it to explore the changing regulatory landscape of ageing. A large variability usually represents a large solution space that is reduced by applied constraints. We have similarly ascertained the variability range for our flux predictions (Figure 4.6B), and can confirm that the inclusion of the transcriptomics data as constraints does reduce the flux variability ranges, in line with what has been previously reported, but inclusion of metabolomics data alongside biomass as the objective function on top of the transcriptomics-derived constraints does not appreciably reduce the flux variability further. This is to be expected, since the method of integration – as an additional component of the objective function – does not aim to constrain the solution space, but rather helps to guide the selection of an optimal solution within the solution space to be more physiologically accurate.

To validate the performance of the novel method, we focused on the TCA cycle because it is conserved between *C. elegans* and mammals, and well annotated in the whole-genome model. We also looked at FEM animals as these metabolically correspond the most closely to wild type without contamination from progeny.

The addition of metabolomics data to the objective function alters the profile of enriched pathways for the comparison between days 5 and 10 in FEM animals. While many of the same pathways as appear in the standard FBA ageing enrichment (Figure 2.7) are still enriched in Metab FBA (Figure 4.7), many new pathway enrichments appear including the expected TCA cycle and oxidative phosphorylation.

The overall sum of fluxes for three indicative pathways across the time series for the two methods is shown in Figure 4.8, illustrating that Metab FBA predicts different fluxes to standard FBA in several aspects of central metabolism.

In particular, the depletion of TCA cycle flux in the later stage of ageing (days 5 to 10) in FEM animals is apparent in Metab FBA but not in standard FBA (i.e. the green square at 257 h in Figure 4.8 is significantly lower than the green square at 137 h). We looked into this prediction in greater detail, as a case study for the effectiveness of the new method.

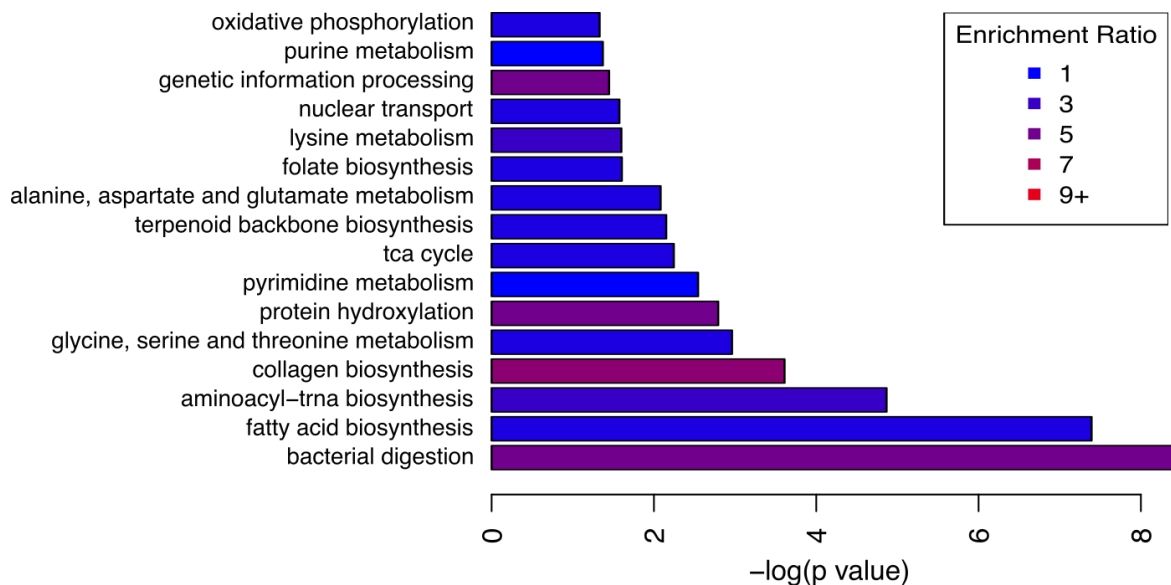


Figure 4.7 Flux enrichment for Metab FBA fluxes, FEM days 5 to 10

Differential pathway flux enrichment for the comparison of predicted fluxes between FEM days 5 and 10 of ageing, showing the pathways whose reactions are significantly changing between days 5 and 10, including the TCA cycle.

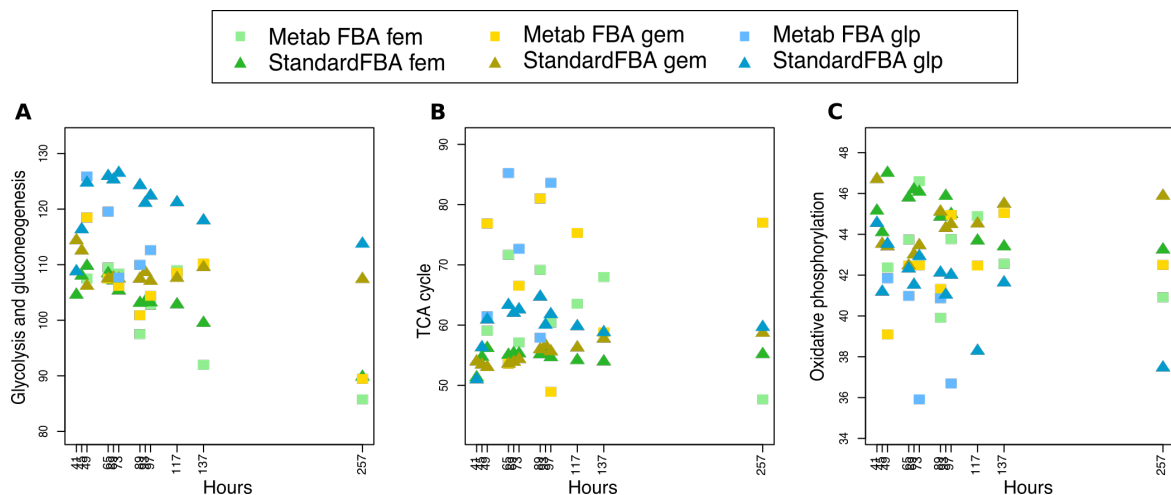


Figure 4.8 Predicted overall fluxes for three pathways

Predicted overall flux sums (the sum of the absolute flux values for all reactions in a pathway) for (A) glycolysis and gluconeogenesis, (B) the TCA cycle, and (C) oxidative phosphorylation, for the two methods: standard FBA (triangles) and Metab FBA (squares). Strains are indicated by colour: blue for GLP, green for FEM, gold for GEM. Metab FBA time points are a subset of the time points available for standard FBA.

As illustrated in Figure 4.9, Metab FBA predicts that fluxes through the reaction MDH_m, which represents the conversion of malate to oxaloacetate by malate dehydrogenase in the mitochondria, was dramatically reduced between the last two time points in

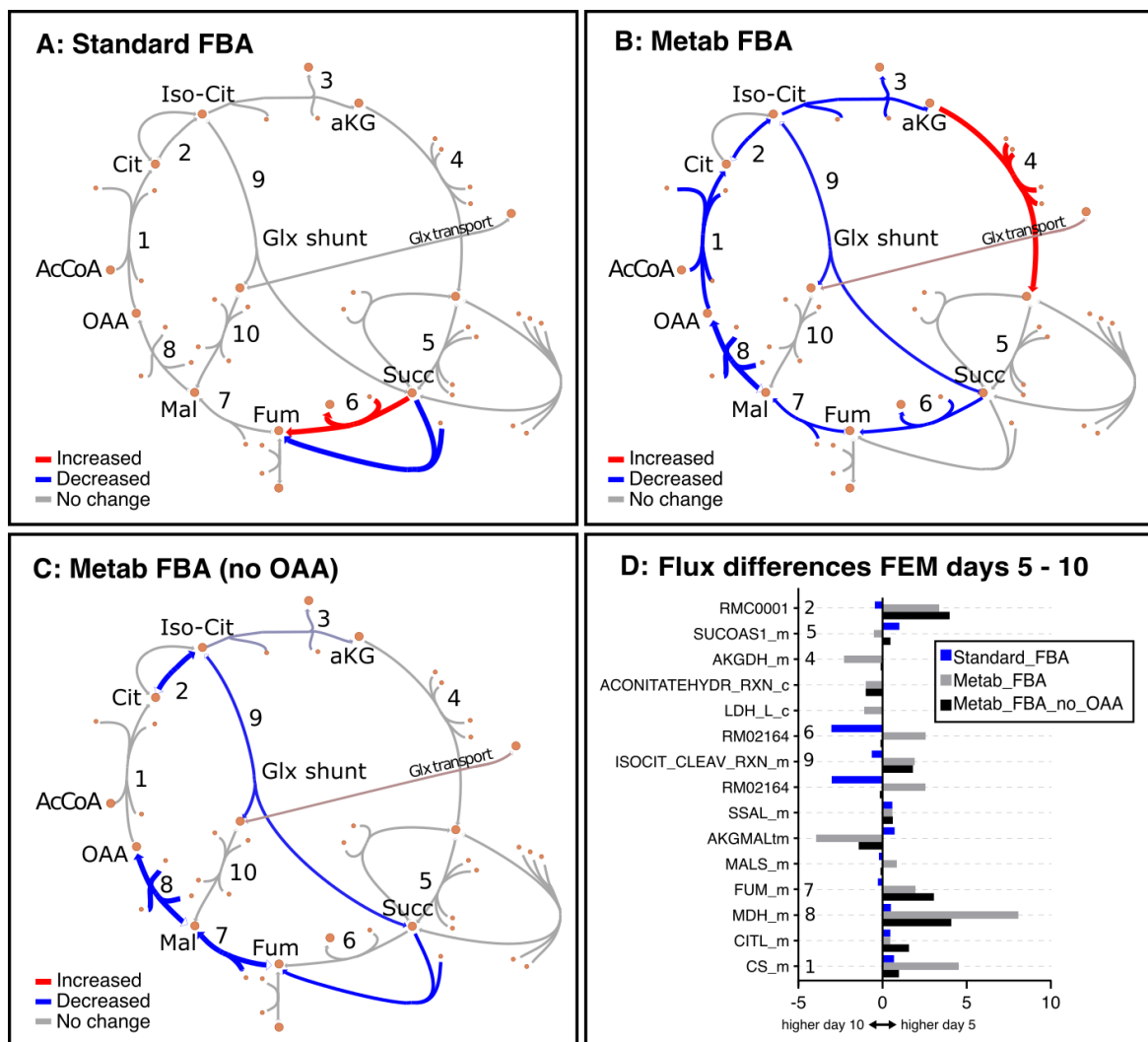


Figure 4.9 TCA cycle fluxes with three different objective functions

(A-C) TCA cycle reactions represented with Escher between days 5 and 10 in FEM animals for standard FBA (A), Metab FBA (B), and Metab FBA (no OAA) (C), which does not include (A) or includes information regarding all quantified metabolites (B) or all except oxaloacetate (C) in their respective objective functions. Red represents increased flux, blue represents decreased flux. (D) The differences between the two time frames are illustrated per reaction for each of the three FBA variants. The reaction identifiers from the WormJam model shown in (D) are as follows: RMC0001: isocitrate hydro-lyase (reaction 1); ACONITATEHYDR_RXN_m: aconitase - isocitrate hydrolase (reaction 1a); RMC0001: Cytosolic isocitrate hydrolase (reaction 2); AKGDH_m: Mitochondrial 2-oxoglutarate dehydrogenase (reaction 3); RXN909_8_m: Mitochondrial succinate dehydrogenase (reaction 4); RM02164: Succinate:ubiquinone oxidoreductase (reaction 5), FUM_m: Mitochondrial fumarate hydratase (reaction 6), MDH_m: Mitochondrial malate dehydrogenase (reaction 7), CITL_m: Citrate lyase (reaction 8), ISOCITDEH_RXN_m: Mitochondrial isocitrate dehydrogenase (reaction 9). This Figure has been adapted from the one published in (Hastings et al., 2019a) (using the January 2018 version of the model), and was originally created by Olivia Casanueva.

FEM. At the same time, CS_m, which represents the conversion of oxaloacetate to citrate in the mitochondria, was also substantially reduced, consistent with the measured reduction of oxaloacetate with age in the metabolomics dataset. As this information was given to the model in the form of the modified objective function incorporating metabolic shifts, one possibility is that the predicted decrease in flux through these reactions is based solely on the direct measurements of oxaloacetate. To validate the *predictive* capability of the model with modified objective function, we re-executed Metab FBA without including the measurement information for oxaloacetate, and called this variant 'Metab FBA (no OAA)'. In this prediction, Figure 4.9C, we observed that fluxes through MHD_m (reaction 1 in Figure 4.9C) remained the most affected within the TCA cycle. This indicates that even without information for this specific metabolite, the model still correctly predicted a substantial drop in oxaloacetate production given the information about the other metabolites.

Considering that if a metabolite becomes limiting with age, then supplementation of the metabolite in the diet should increase survival, or in other words, extend lifespan, it is interesting to note that while supplementation of most TCA metabolites either does not affect lifespan, or has only a small effect (Edwards et al., 2015, 2013), oxaloacetate extends lifespan by 25% and it depends on both AMP-activated protein kinase (AMPK) and insulin signaling (Williams et al., 2009). Oxaloacetate supplementation elevates the levels of NAD⁺ and restores redox balance, acting through sirtuins and AMPK (Roth and Ingram, 2016). In mouse models of stroke, oxaloacetate administration has been reported to reduce neural damage and traumatic brain injury (Roth and Ingram, 2016). The fact that oxaloacetate supplementation has been shown to have such an important impact on lifespan also indicates that dysfunctional mitochondria are not just a co-morbidity related to age deterioration, but rather that this is one of the drivers of the aging process. Moreover, the predicted flux changes due to age bear a resemblance to flux changes due to explicit models of TCA dysfunction, such as fluxes reported for an *idh-1* knock-out mutant (Vergano et al., 2014), illustrated in Supplementary Figure A.7 in the Appendix.

Taken together, these studies illustrate how the composite objective function in Metab FBA outperforms a standard objective function based on growth alone by providing more accurate predictions.

Limitations

The most significant limitation of the Metab FBA approach is that it relies on an underlying comparison of the metabolomics concentration values between neighbouring time points, which means that the resulting flux predictions are applicable, strictly speaking, to the time frame that represents the comparison between those two original time points. Thus, for

comparisons outside of the sequential progression of time (e.g. between strains, or between non-sequential time points) it is not possible to interpret fluxes predicted with this method.

Another important limitation of this approach is that incorporation of the metabolomics differences in the objective function without quantitative constraints may result in the model having too much freedom to optimize production or consumption of those specific metabolites, unbalancing the overall metabolic flux away from realistic levels. It is apparent that the Metab FBA predictions in Figure 4.8 show a wider range of variation than the standard FBA predictions – e.g. in Figure 4.8B, the TCA cycle flux predictions, both the highest and the lowest yellow squares that represent Metab FBA GEM total fluxes at different time points, are *outside* of the range of the yellow-ish triangles that represent standard FBA total fluxes at different time points. Later, I will argue that this variability in part derives from greater biological accuracy – the metabolic fluctuations with time of day, for example, that were described in Chapter 2, are more apparent in Metab FBA than standard FBA – but the metabolomics integration only provides information of the *direction* of change rather than the *magnitude* of change. To mitigate against unconstrained optimisation of flux through the added metabolomics-derived boundary reactions, an arbitrary maximum threshold has been incorporated for boundary fluxes implementing the metabolomics integration.

Of course, it would be better in the future to add a quantitative element to the metabolomics data integration with the flux prediction. Moreover, the use of neighbouring time points rather than overall trends in the metabolomics time series measurements to calculate differences for incorporation into the model may amplify noise. For both of these reasons, it would be desirable to calculate a linear rate of change for metabolite level changes across a number of time points (as is done, for example, in Bordbar et al. (2017)) and use those per-metabolite linear rates of change as constraints on the relevant reactions that are added to the objective function. However, this enhancement has not been implemented and is indicated for future work.

The analysis that follows is mindful of these possible pitfalls. In particular, we limit the use of Metab FBA to pairwise comparisons between neighbouring time points, thus falling back to standard FBA for interpreting the overall metrics.

4.3.2 Overall growth rate and environmental exchanges (Standard FBA)

Figure 4.10A shows the optimal (maximized) growth rate achieved for standard FBA for each of the strains and time points. Growth rate – flux through the biomass reaction – in post-mitotic adults can be thought of as a proxy for metabolic rate, i.e. the turnover of constituent energy sources in order to provide fuel for the maintenance of an organism.

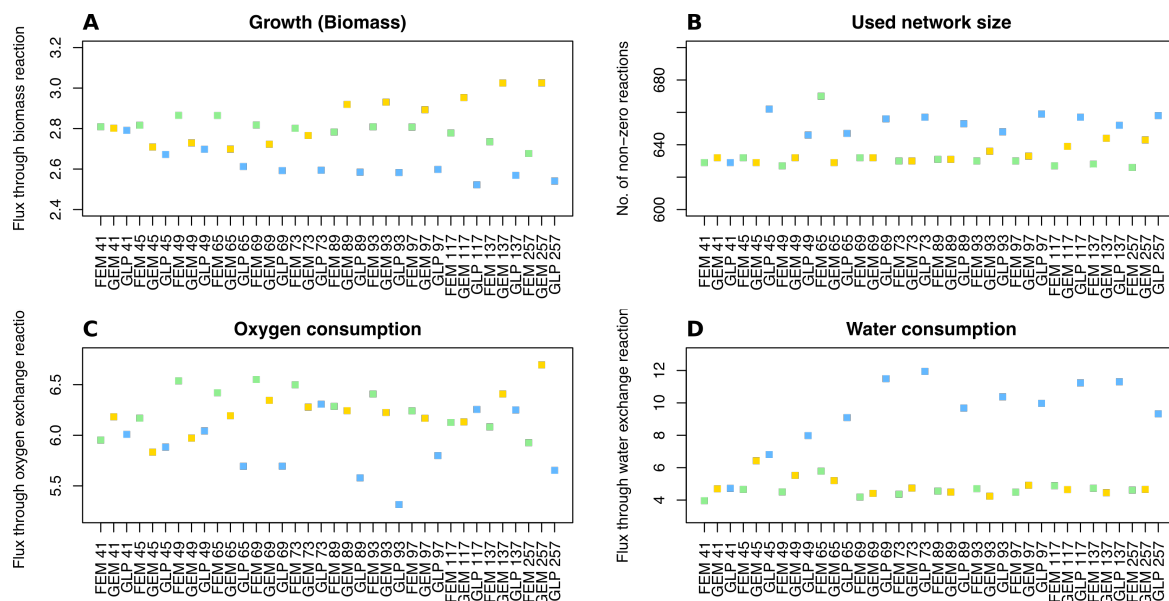


Figure 4.10 Overall metrics for the growth of the organism

The Figure shows important indicators selected from the standard FBA predicted fluxes across all strains and time points. (A) Growth rate: flux through the biomass reaction. (B) Network usage: total number of reactions with non-zero flux in the solution. (C) Oxygen consumption: uptake of oxygen into the model. (D) Water consumption: uptake of water into the model. Strains are indicated by colour: blue for GLP, green for FEM and gold for GEM.

It is apparent that, broadly speaking, the GLP strain is consistently hypometabolic (lower growth rate) compared to the FEM and GEM strains, i.e. in Figure 4.10A the GLP (blue) values are lower than the other two strains. Moreover, the metabolic rate of GEM animals appears to be increasing with age in a way that may correspond to the contamination of the samples with progeny related to the incomplete penetrance of the mutation, while the FEM strain, which can be considered closest to wild type metabolically, shows a gradual decline in maximal growth rate with the course of ageing.

Interestingly, while the GLP shows on average lower growth rates, it is using more of the network to achieve that: the total network usage is consistently higher in that strain (Figure 4.10B). Oxygen consumption is predicted to be lower in GLP animals at most time points, corresponding to lower rates of oxidative phosphorylation, although not consistently and also not linearly with respect to time (Figure 4.10C). Intriguingly, the GLP strain is predicted to require greater consumption of water during ageing (Figure 4.10D), hinting at quite different metabolic processing taking place in that strain.

Still using standard FBA, I explored the exchange fluxes more generally (Figure 4.11). Interestingly, the GLP strain is the only strain that shows regular and active use of the glyoxylate shunt pathway (Figure 4.11A), indicating a differential pattern of energy usage compared to wild type. As discussed earlier, GLP animals are known to be fatty relative to

wild types, and their altered pattern of fatty acid metabolism has been shown to be related to their longevity.

The model also predicts that GLP animals systematically excrete more acetate than the other two strains (Figure 4.11B), derived from an excess of acetyl-coA relating to the activation of energy sources from lipids, even though we do not specifically provide information about lipids to standard FBA, and the fat-related pathways are not well annotated in the model.

GLP animals also consistently excrete more ribose than GEM and FEM (a sugar; Figure 4.11C), and they are the only strain to consistently excrete guanine (a nucleobase, Figure 4.11D), although they only start this at day 2 (65 h). Moreover a pattern of altered nucleotide metabolism is clearly reflected in the shift between days one and two (49 - 65 h) from xanthine excretion towards guanine excretion (Figure 4.11F). By contrast, hypoxanthine (a known longevity modulator) excretion is absent in GLP while apparent in the other two strains (Figure 4.11E). Taken together, these differences reflect an increased reliance on nucleotide and fatty acid metabolism for energy in GLP.

Supplementary Figure A.8 in the Appendix illustrates the comparison of fluxes between FEM and GLP at 49h, the time point at which the GLP metabolism starts to shift noticeably away from the other strains according to the overall indicators. It is apparent that many of the reactions in the TCA cycle are notably different, including the predicted switch-on of the glyoxylate shunt pathway in GLP which is not seen in the other strains, coupled to an increase in electron transport chain complex I. Malate dismutation is also more active in GLP, and we see differences in amino acid production and consumption rates as well as in nucleotide metabolism.

In addition, breakdown of malonyl-CoA and acetyl-CoA is seen to generate higher rates of acetate and acetaldehyde. The predicted export of acetate (and also acetaldehyde – data not shown) hints at a potential connection to epigenetic histone modifications. We can speculate that the greater availability of acetyl-CoA that is driving the export of acetate in the model might, *in vivo*, facilitate a greater level of donation of acetyl groups to histone acetylation processes, leading to overall higher levels of acetylation.

4.3.3 Age-related changes in pathway fluxes (Metab FBA)

Using Metab FBA and differential pathway flux enrichment, we can explore the sequence of altered flux profiles between each consecutive time point in the time series, in order to better understand the metabolic changes that take place during ageing. Metab FBA has fewer time points than standard FBA because our metabolomics dataset contained fewer time points than our transcriptomics dataset (i.e. only a subset of samples were sent for metabolomics). Moreover, because Metab FBA works on the basis of a comparison of metabolomics

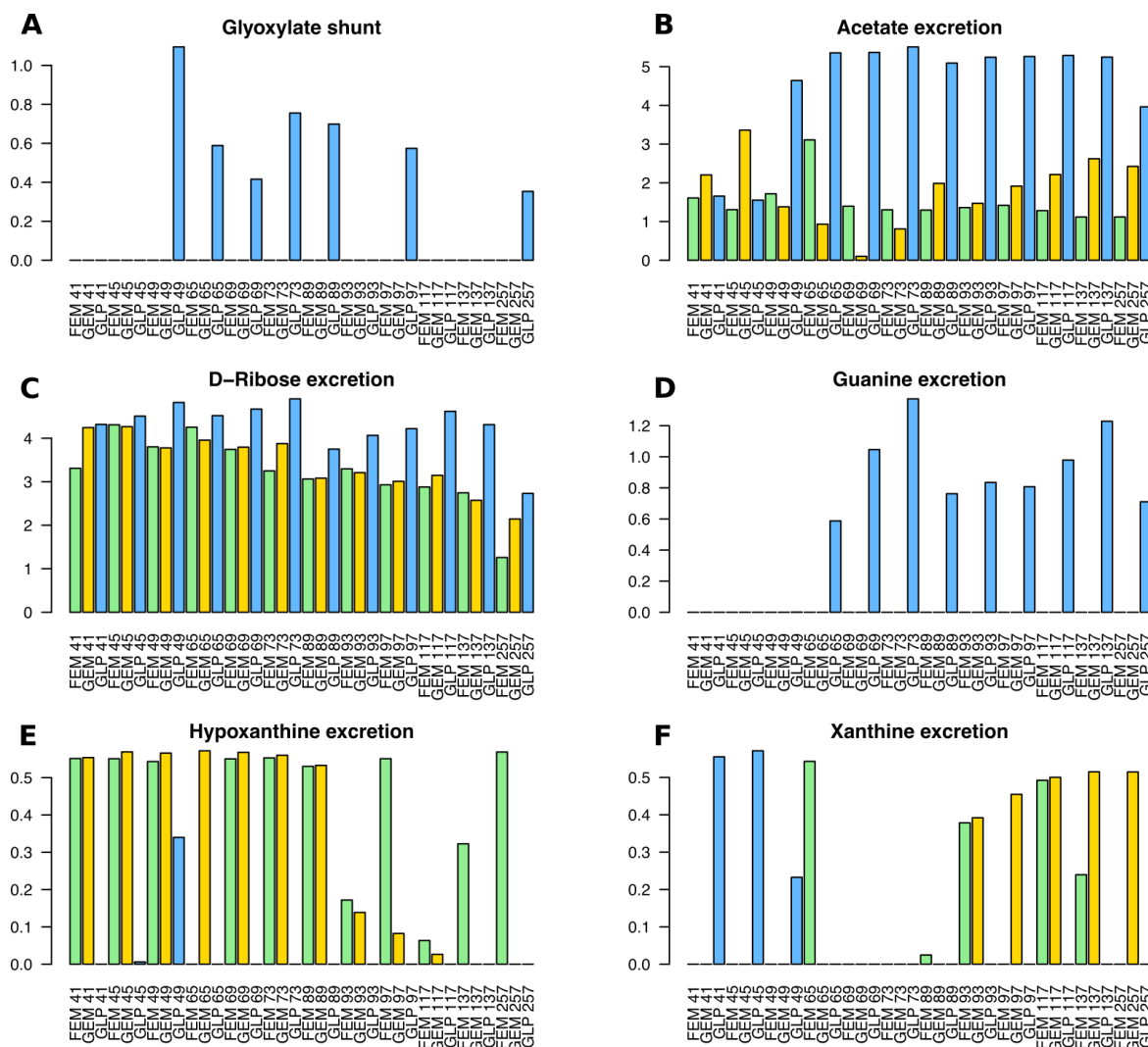


Figure 4.11 Fluxes through key reactions indicating energy usage

The Figure shows standard FBA predicted fluxes through key reactions in the metabolic network related to energy usage and by-product excretion. (A) Glyoxylate shunt: fluxes through the isocitrate cleavage to glyoxylate reaction. (B) Acetate excretion: flux through acetate excretion at the system boundary. (C) D-Ribose excretion: flux through ribose excretion at the system boundary. (D) Guanine excretion: flux through ribose excretion at the system boundary. (E) Hypoxanthine excretion: flux through hypoxanthine excretion at the system boundary. (F) Xanthine excretion: flux through xanthine excretion at the system boundary. Strains are indicated by colour: blue for GLP, green for FEM and gold for GEM.

measurements at neighbouring time points, the first metabolomics measurement is not able to be used predictively.

Normal *C. elegans* ageing – FEM animals

For reasons described earlier, we consider the FEM strain the most wild type-like in our study and thus use it as a proxy for normal *C. elegans* ageing.

To summarise the overall time series, Figure 4.12 gives the enriched pathways for each of the neighbouring comparisons between consecutive time points in the time series, indicating which pathways have a significant number of reactions within them that can be considered meaningfully different between those time points.

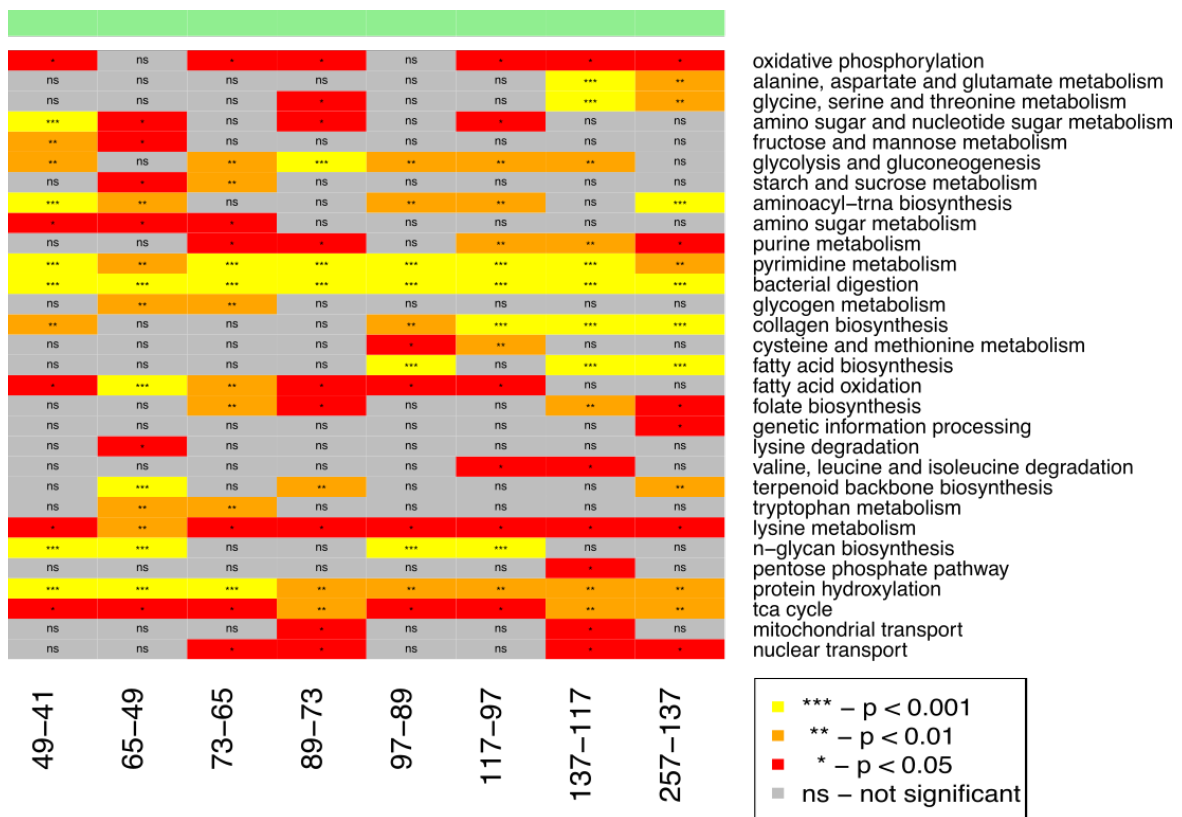


Figure 4.12 Pathways with flux differences between consecutive time points in FEM

Pathways enriched for flux differences between consecutive time points across the full time series for FEM animals. Each column represents a comparison between two consecutive time points. As there are no Metab FBA flux predictions at time 41 h, the 49-41 h comparison in the table shows the comparison of 49 h Metab FBA fluxes with 41 h standard FBA fluxes. A pathway being different between two time points means that more reactions in that pathway had different fluxes than would be expected by chance alone, but does not say which direction the changing flux was in.

According to these predictions, some pathways, such as oxidative phosphorylation, are affected both during early ageing and later ageing. Others, such as alanine, aspartate and glutamate metabolism, are only affected during later ageing, while amino sugar metabolism is only changing during early ageing. In general, each pathway has a slightly different

profile with respect to temporal changes, reflecting that each time point a slightly different metabolic subsystem is predicted to be optimal given the multi-omics dataset that is framing the search space. Moreover, Figure 4.12 is suggestive that there are shifts back and forth during the earlier time points. However, the pattern of enrichment for between-time-point changes in predicted pathway fluxes does not give any information about whether flux is increasing or decreasing, or indeed exactly which reactions are affected. To understand the flux enrichments better, it is necessary to dig deeper into the predicted fluxes.

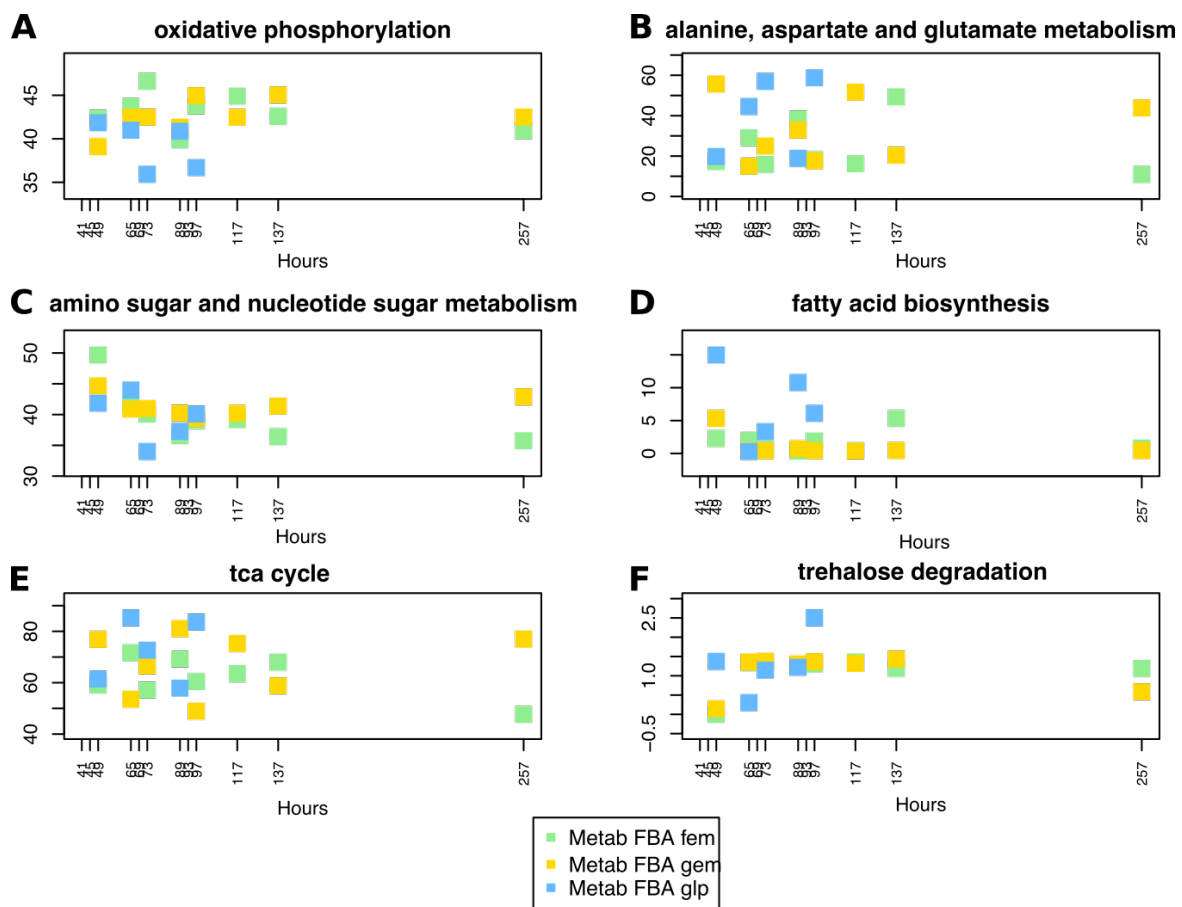


Figure 4.13 Metab FBA total sum of fluxes predicted for selected pathways

The figure shows the sum of the absolute value of predicted fluxes for reactions belonging to selected pathways: (A) oxidative phosphorylation; (B) alanine, aspartate and glutamate metabolism; (C) amino sugar and nucleotide sugar metabolism; (D) fatty acid biosynthesis; (E) TCA cycle; (F) collagen biosynthesis. Colours indicate strain: blue for GLP, green for FEM, and gold for GEM.

Figure 4.13 shows the overall sum of absolute predicted flux values per pathway for a selected subset of the pathways that are enriched for age-related changes between at least one pair of neighbouring time points. Some of the changes due to later ageing (between hours 137 and 257, i.e. days 5 and 10 in our time series) have already been discussed above, e.g. the decrease in flux through the TCA cycle. Between days 5 and 10, there is also a

decrease in flux through oxidative phosphorylation (Figure 4.13A), alanine, aspartate and glutamate metabolism (Figure 4.13B), and fatty acid biosynthesis (Figure 4.13D).

It is apparent that many of the pathways show relatively large variability in flux values during the early ageing time points. Some of the flux changes between earlier time points appear rhythmic, indicating they may be related to possible time of day effects which may correspond to the observed influence of circadian changes in the metabolome discussed in Chapter 2. In FEM animals, this is most pronounced in the TCA cycle (Figure 4.13E), having the highest predicted flux during the earlier (morning) time points, corresponding to when the animals would be more active. A similar pattern is observed in alanine, aspartate and glutamate metabolism (Figure 4.13B), again linking the TCA cycle to flux through these amino acids. By contrast, in FEM animals, predicted fluxes through fatty acid biosynthesis (Figure 4.13D) are relatively stable throughout the time series, and amino sugar and nucleotide sugar metabolism (Figure 4.13C) shows a non-rhythmic steady decrease over the full course of the ageing time series.

Although we do not discuss the GEM strain predicted fluxes in detail, it is also apparent that some strain differences are predicted. In particular, the GLP animals (for which we do not, unfortunately, have metabolomics measurements at the later ageing time points, thus are not able to use the Metab FBA during later ageing) show a quite different pattern in these selected pathways during early ageing, which we discuss next.

Early ageing in a longevity model – GLP animals

Based on Figure 4.13, we can make several observations about the predicted differences in GLP metabolism during early ageing relative to our model of normal ageing, FEM. Oxidative phosphorylation is predicted to be somewhat lower in GLP than FEM, and interestingly is predicted to vary rhythmically in GLP but not noticeably in FEM, decreasing noticeably in GLP during the evening time points. By contrast, fatty acid biosynthesis (Figure 4.13D) is predicted to be generally higher in GLP animals, in particular during the evening time points. This is of course consistent with our observation in the metabolomics dataset (discussed in Chapter 2) that the GLP animals appear to be more fatty in the evenings. Alanine, aspartate and glutamate metabolism (Figure 4.13B) is also predicted to be elevated in GLPs in the evenings. Interestingly, trehalose *degradation* is predicted to steadily rise in GLPs during the course of early ageing (Figure 4.13F), while staying relatively low and constant in the other two strains. Trehalose is a pro-longevity storage molecule (Honda et al., 2010). Blocking an age-related shift from storage of excess energy as trehalose (which occurs early in ageing) to storage as glycogen (which occurs later in ageing), i.e. storing excess sugars as trehalose rather than glycogen for longer, has also been shown to be pro-longevity in *C. elegans* (Seo et al., 2018).

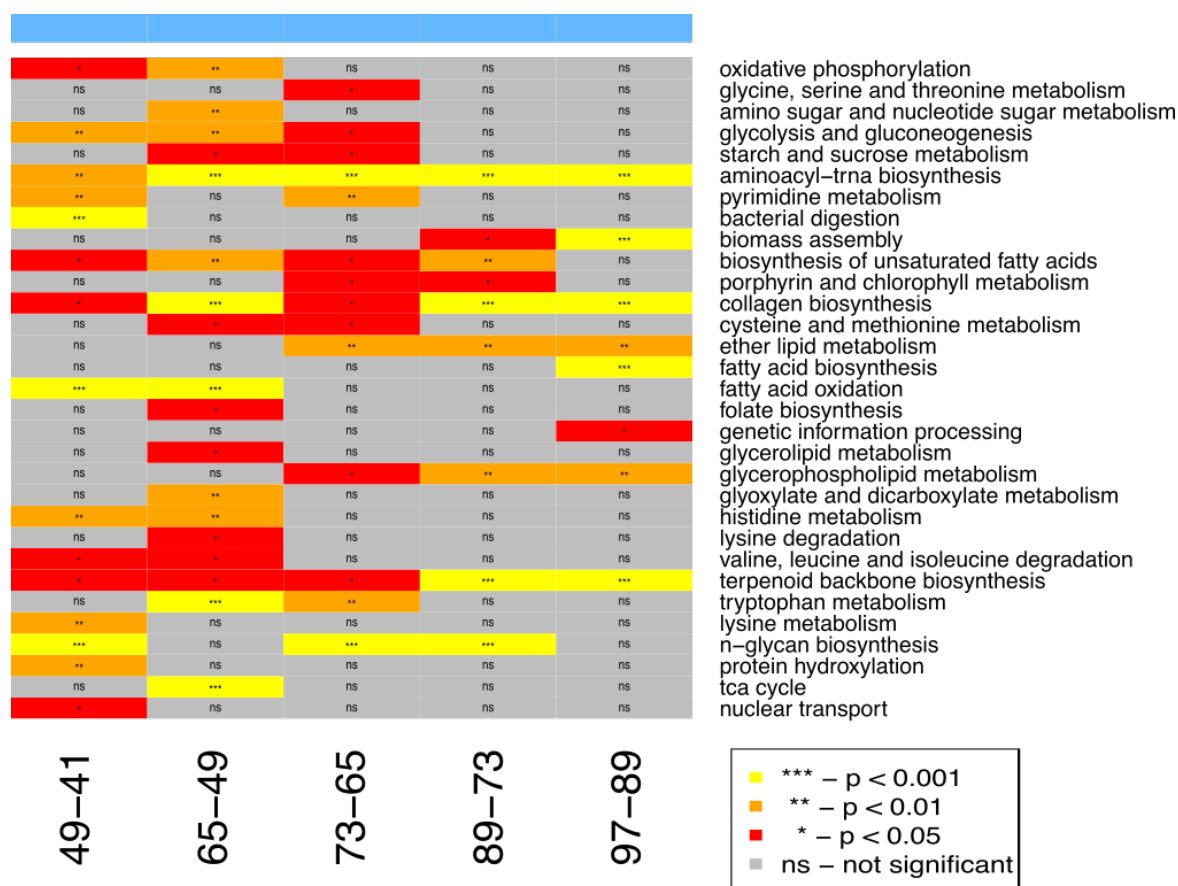


Figure 4.14 Pathways with flux differences between consecutive time points in GLP

Pathways enriched for flux differences between consecutive time points across the full time series for GLP animals. Each column represents a comparison between two consecutive time points. As there are no Metab FBA flux predictions at time 41 h, the 49-41 h comparison in the table shows the comparison of 49 h Metab FBA fluxes with 41 h standard FBA fluxes. A pathway being different between two time points means that more reactions in that pathway had different fluxes than would be expected by chance alone, but does not say which direction the changing flux was in.

Looking at the enrichment for changing fluxes in pathways for GLP animals (Figure 4.14), it is apparent that many pathways are shifting between day 1 and 2 (time points 49 and 65), including glycerolipid metabolism, pyrimidine metabolism, lysine metabolism, glyoxylate metabolism and n-glycan biosynthesis.

We can explore some of these pathways further in Escher, as illustrated in Supplementary Figure A.9 in the Appendix, which shows individual predicted reaction flux discrepancies in multiple reactions in glycolysis and the pentose phosphate pathways, connected to increased trehalose production. Moreover, some of the most different reactions in the network relate to fatty acyl CoA metabolism, as indicated for example for palmitoyl-CoA and linoleoyl-CoA, indicating greater energy metabolism through fats, as would be expected for the strain at that time point, which is the timing at which GLP animals undergo a

fundamental metabolic shift and gain their pro-longevity metabolic patterning (Mains, 2018).

As mentioned above, it is intriguing that the GLP animals, while being ‘hypometabolic’ with respect to growth rate (i.e. having a predicted lower biomass production rate), are predicted to be using more of the available reaction network to achieve that growth (i.e., more metabolic reactions are active; Figure 4.10). This is even more the case in the Metab FBA prediction than in the standard FBA prediction (Figure 4.15).

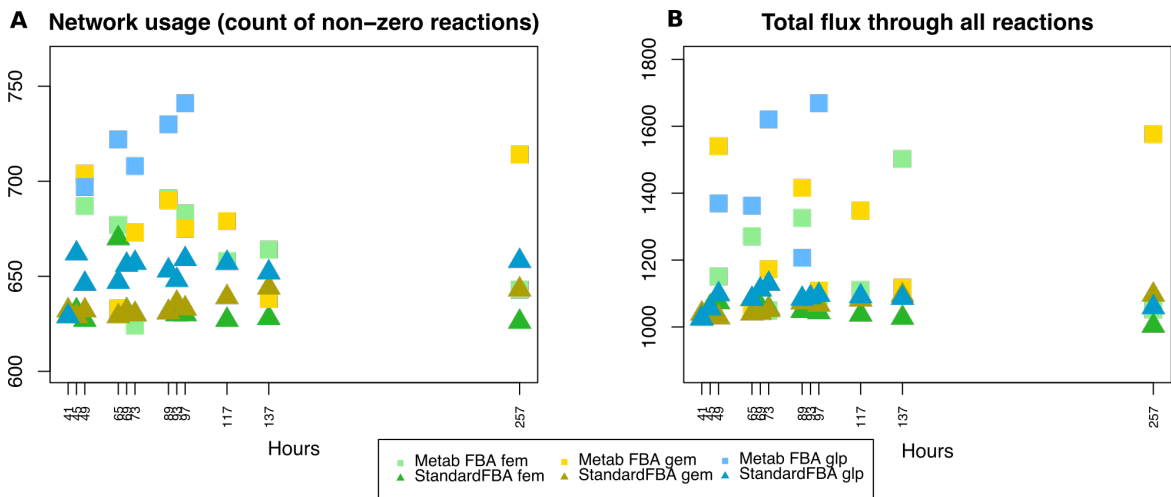


Figure 4.15 Predicted reaction network usage and total flux sum

(A) Network usage: the count of non-zero reactions within the metabolic network in a given flux solution. (B) Total flux sum: the sum of the absolute value of predicted fluxes through all reactions in the network.

If accurate, this would be consistent with what is known about the germline-free longevity phenotype, i.e. that it is dependent on a global rewiring of the metabolic network, and that it is fragile in the sense that the GLP longevity is dependent on many otherwise unrelated aspects of metabolism (Mains, 2018).

4.4 Conclusion

Overall, we observe that the use of FBA together with a multi-omics dataset can provide intriguing predictions with respect to different patterns of metabolism during ageing. FBA is a well established method in unicellular organisms such as *E. coli* and yeast, but has only recently been available for *C. elegans*.

We developed a novel approach to harness metabolomics measurements together with transcriptomics measurements in order to improve the accuracy of FBA predictions. We validated our method through evaluation of a case study of the TCA cycle in ageing FEM

animals, for which we were able to demonstrate that our method is able to predict changes in metabolite levels for which no metabolite information is provided, thus with the addition of relatively easily obtainable data (e.g. metabolites for which good standards exist), our model may be used to predict system level metabolic shifts or changes in metabolites that are more difficult to quantify (i.e. those without appropriate standards).

We used this integrative method to probe the changes due to ageing during our time series. During later ageing the model is able to reproduce several known aspects of ageing *C. elegans* physiology, such as a decline in oxidative phosphorylation and TCA cycle. During early ageing, the predicted fluxes further illuminate the rhythmic shifts that take place in the metabolome during early ageing. Several aspects of metabolism differ between the normal-lived strains and the GLP longevity model, and these are also apparent in our predicted fluxes: GLPs have higher levels of fatty acid metabolism in various aspects and altered processing of nucleotides and waste products such as acetate.

There are several limitations to the work presented in this chapter. Assessing and interpreting predicted fluxes in a large-scale model such as this is inherently challenging, and the methods that have been used for the interpretation of predicted fluxes – differential flux analysis, flux pathway enrichment, total flux sums for pathways, visualisation of individual reaction fluxes with Escher pathway maps – all suffer from limitations and offer only a partial view on the underlying, hugely complex, genome-scale flux predictions. Various thresholds are used in these analyses which are to some extent arbitrary, in the selection of which I have had to be pragmatic. More research is needed in these areas to make FBA better applicable to be used in the context of large-scale multi-omics time series experiments.

The flux predictions may be noisy, both due to redundancy (alternative optimal pathways) in the underlying network model, and due to the method of integrating the metabolomics data, which only provides a guide as to the direction of fluxes and not their magnitude. Moreover, as discussed in the previous chapter, the underlying model is still incomplete in several areas, and even of the network that has been annotated, it is far from fully used in the flux predictions. There are several pathways that are not used at all by any of the the flux solutions and which would be very interesting if they were, e.g. various epigenetic modification pathways such as histone acetylation. At present, within the steady-state solution optimising both biomass production and metabolomics changes, there is nevertheless no incentive for the model to use the histone modification pathways. With an appropriate integration method, it would be ideal to add yet another -omics layer to give appropriate information about this aspect of biology to the predictive model.

While it may be difficult to interpret the predicted fluxes, the wealth of information provided by FBA is so rich as to yield many and diverse insights when probed in different ways. In this chapter I have looked at the fluxes themselves and their evolution over time

in the context of pathways. In the next chapter, among other network-based approaches to interpret the dataset as a whole, I look at some properties of the flux balance problem as a whole, in terms of how flexible and how constrained it is in each of the different conditions.

5. Metabolic regulation of ageing

5.1 Introduction

Thus far, the previous chapters have reported observations about the changes in metabolism taking place during ageing. However, the key question that arises in pursuit of greater health across the lifespan is *How can we intervene to promote health for longer?* Ideally, interventions in the ageing process to promote health would take the form of small molecules that can be developed towards treatments. To this end, in this chapter I seek to go one step beyond the observational studies reported in previous chapters, and investigate our dataset for evidence of novel modulators of the ageing process.

Firstly, I explore the flux variability and sensitivity parameters associated with the network of predicted fluxes arising from the time series of FBA flux predictions to identify whether the flexibility or constraints of the network change over time.

Secondly, considering only the metabolomics dataset, I look at how the measured metabolites cluster together with respect to their concentration level changes over time. Those metabolites that behave similarly to known longevity modulators are good candidates to be longevity modulators themselves. I introduce a novel network-based ranking for metabolites based on their similarity to known longevity modulators.

Finally, I harness the linked transcriptomics dataset and consider a trans-omics approach relating genes to metabolites based on a joint correlation analysis. The resulting bipartite network can inform enquiries across the levels and uncover novel aspects of metabolic regulation during ageing.

5.2 Methods

5.2.1 Regulation and Bottlenecks in Flux Predictions

Flux Variability Analysis

As the FBA problem is underconstrained, each optimal solution is in principle compatible with a range of flux values in all but the most constrained reactions. To assess the possible range of flux values for each reaction compatible with the same optimal solution, the variability of flux through each reaction can be explored in a multi-step linear optimization known as flux variability analysis (FVA). This approach alternately maximises and minimises flux through each reaction one by one in order to determine the maximum and

minimum possible flux for each reaction compatible with the same objective function value (Burgard et al., 2001).

In my pipeline, flux variabilities are computed using the CobraPy 'cobra.flux_analysis.flux_variability_analysis' method, as a second step after obtaining the optimal value for the objective function. Flux variability gives a maximum and a minimum flux for each reaction; to obtain the flux variability range I calculated (maximum - minimum). Constraints and objective function corresponding to standard FBA as described in the previous chapter were used in this analysis, thus the full time series of data including in GLP animals can be presented.

Within the predicted flux variabilities, there were two outliers: time points 117 and 257 in the GLP strain. These groups consistently show a far greater flux variability than the average of all the other groups, without having any other significantly different parameters (growth rate, metabolic inputs etc.). The most likely reason for this discrepancy is technical. For example, there may be an undiagnosed flux loop that is enabled only in those cases and not in the others. These two GLP time points were therefore regarded as outliers and ignored in this analysis.

Reduced costs for reactions and shadow costs for metabolites

Reduced costs and shadow prices are measures of sensitivity that are returned from the CobraPy implementation that solves the FBA problem. Shadow prices of metabolites represent the sensitivity of the objective function value (that is being maximised or minimised) to changes in the availability of those metabolites. The value given for the shadow price can be negative, zero, or positive, depending on the intrinsic value of the metabolite. When the objective value is being maximised, the most constrained metabolites will be assigned negative shadow prices.

Similarly, reduced costs of reactions represent the sensitivity of the objective function value to additional flux through a particular reaction. Reduced costs of reactions can also be negative, zero or positive, and similarly, when the objective value is being maximised, the most constrained reactions will have negative reduced costs.

These metrics are interrelated, in that flux limiting metabolites will lead to reactions involving those metabolites being similarly constrained. For our analysis in this chapter, we focus mainly on the reduced costs of the reactions rather than the shadow prices of the metabolites, as the reactions are annotated to pathways which can be used for interpretation. For interpretation I have used a variant of the pathway enrichment analysis that was introduced in the previous chapter, again using Fisher's exact test, but this time rather than comparing two conditions, I have used the set of reactions with reduced costs < 0 for a given time point and strain, with the whole set of model reactions as the background.

This gives a predicted enrichment (over-representation) statistic for each pathway for each time point and strain.

5.2.2 Clustering and network inference

Metabolite network inference

A network was inferred based on standard Pearson correlations between metabolites across all replicates within the metabolomics study. A threshold of 0.65 in absolute strength of correlation, and an adjusted p-value of 0.05 was used to define an edge between two metabolites being present in the network. Correlations were calculated and assigned p-values with the 'psych' library (Revelle, 2018) function 'corr.test' in R. The resulting correlation network was visualised using Cytoscape (Su et al., 2014).

Gene-metabolite network inference

Correlation analysis of measured values between genes and metabolites can discover clusters of co-regulated genes and metabolites that vary together across conditions and samples, as was done for the set of transcripts and metabolites found in human blood in (Bartel et al., 2015). Statistical whole-systems approaches along these lines can also be used to make predictions about causal relations (Rosato et al., 2018).

For the gene-metabolite network inference I present in this chapter, the metabolomics dataset was combined with a subset of the transcriptomics dataset that was taken from the same samples across strains and time points. A network was inferred based on standard Pearson correlations between genes and metabolites across all replicates. Only positive correlations were considered, as negative correlations were more numerous and may be harder to interpret. A threshold of 0.7 in correlation strength was applied, and an adjusted p-value of 0.05 was used to define an edge being present in the network. Correlations were calculated and assigned p-values with the 'psych' library (Revelle, 2018) function 'corr.test' in R. As for the metabolite-metabolite network, the resulting correlation network was visualised using Cytoscape (Su et al., 2014).

5.3 Results and Discussion

5.3.1 Flexibility and constraint in the metabolic network

FBA predicts reaction fluxes through metabolic reactions in a constrained overall system under the assumption of a steady state. Alongside predicted fluxes, FBA offers sensitivity parameters with each solution, which indicate which parts of the overall network are

constrained in that solution, as compared to which are free to vary around a range. These parameters can then be compared between different time points to understand how flexibility is changing during the ageing process.

Flux Variability Analysis

FVA gives a prediction of the flux variability (FV) range compatible with a given optimal solution, indicating which reactions are more constrained and which have more flexibility. I calculated the FV range for every reaction in the network for each of the time points and strains. I then amassed these data into sums by pathway for broad interpretation. The overall differences in FV range and a selection of pathways are indicated in Figure 5.1.

It is apparent in Figure 5.1A that the overall sum of flux variability ranges indicates what we might expect with regard to strain differences in overall temporal profile: the FEM strain starts off with the largest flexibility but then decreases, gradually during early ageing and more sharply between time points 137 and 257 at the end of the time series, while GEM appears to increase somewhat during middle age, possibly corresponding to the contamination due to incomplete penetrance in that strain, and GLP has lower metabolic flexibility on average but also maintains a more stable trajectory in later ageing.

The TCA cycle (Figure 5.1B) is perhaps one of the pathways that shows the effect of ageing the most: there is a gradual depletion of FV range with age in this pathway across all strains, showing that the system overall has less flexibility in key aspects of energy generation as it ages. Intriguingly, one of the pathways that shows the greatest discrepancy between the strains is nuclear transport (Figure 5.1C): the movement of metabolites such as S-adenosyl-methionine and NAD, involved for example in histone modifications, into and out of the nucleus. The variability range of these reactions are significantly depleted in GLP as compared to the other strains, and shows an ageing-characteristic depletion with age in GLP and FEM, but not the GEM strain. In fact, in GEM the variability range in nuclear transport increases steadily throughout the time series, indicating that the GEM incomplete mutation penetrance may be affecting this pathway, which may be expected as if a subset of GEM animals are fertile, these animals would be undergoing active cell division.

Mitochondrial transport (Figure 5.1D) shows a similar pattern to nuclear transport, although less pronounced, which interestingly is also very similar to the pattern for fatty acid metabolism (Figure 5.1F), indicating that the transport of metabolites (such as acetate and carnitine) into and out of the mitochondrion is a rate limiting factor for fatty acid metabolism, as might be expected. Flux variability in fatty acid metabolism is lower in GLP animals, indicating that this strain is more dependent on fatty acid metabolism for achieving normal growth, as might be expected based on what is known about the GLP phenotype.

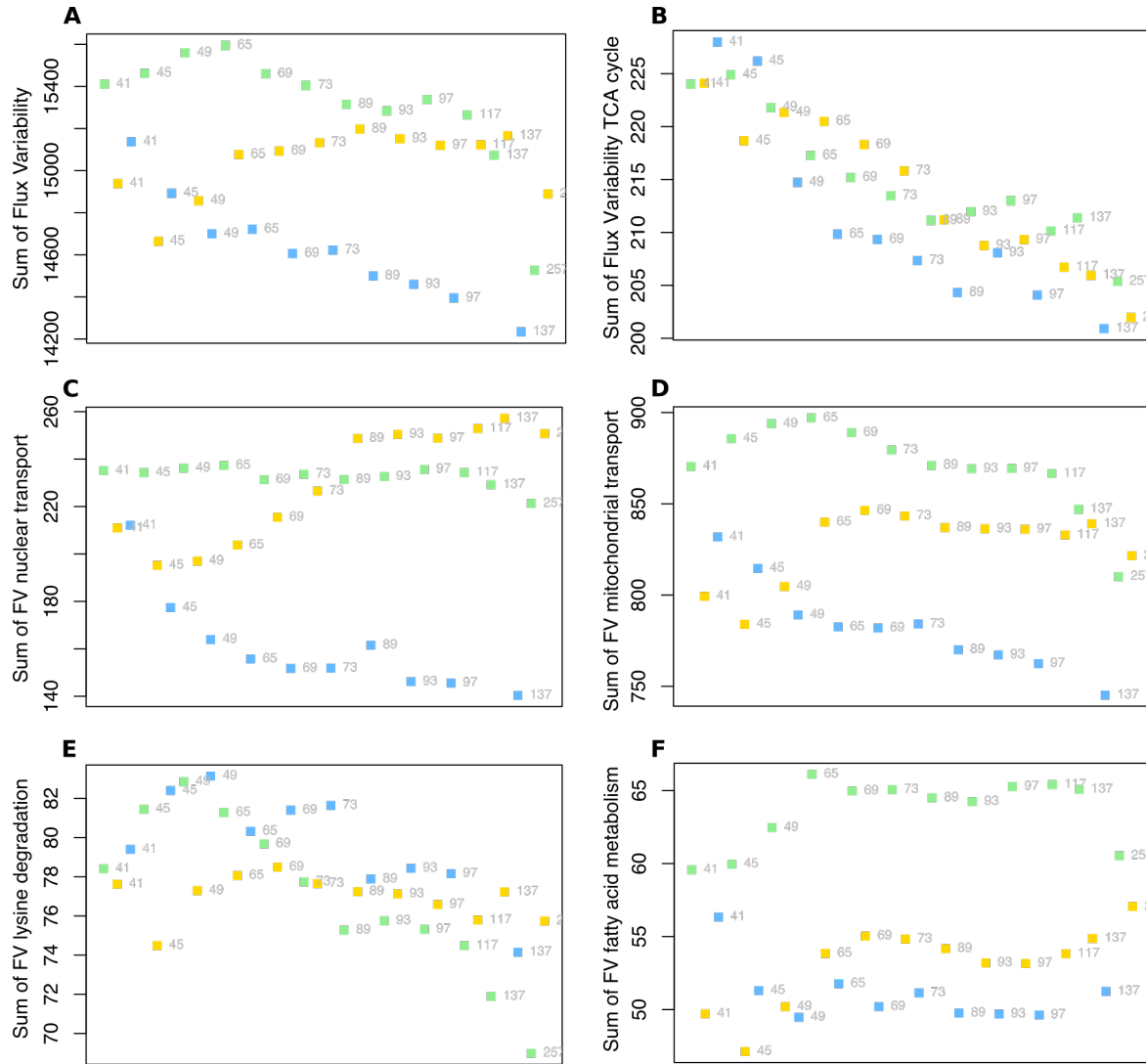


Figure 5.1 Flux Variability range sums per pathway

The Figure shows the Flux Variability (FV) ranges summed up for (A) the whole network, and (B-F) specific pathway subsets of reactions within the network: (B) TCA cycle; (C) Nuclear transport; (D) Mitochondrial transport; (E) Lysine degradation; (F) Fatty acid metabolism. In each plot, time points are indicated by labels and strain by colours: blue for GLP, green for FEM and gold for GEM. The GLP FV range at time points 117 and 257 have been removed as they were outliers. Constraints corresponding to standard FBA were used in this analysis.

An example of a pathway that shows a sharp drop in FV range with age across all strains is lysine degradation (Figure 5.1E), a pathway that produces energy intermediates such as *S*-adenosyl-methionine from amino acid precursors. In mammals, lysine degradation has been implicated as important for promoting healthy brains over the life course (Hallen et al., 2013).

Reduced costs and shadow prices

The *reduced cost* of a reaction represents the sensitivity of the objective function value (optimal predicted outcome flux) to changes in the flux value of that reaction. That is, reduced costs of reactions capture the dependence of the final optimized solution value on a particular reaction value. *Shadow costs* of metabolites, on the other hand, are sensitivities to availability in a given metabolite (Maarleveld et al., 2013) – that is, they track the dependence of the final optimized solution value on availability of a particular metabolite. Shadow prices have been shown experimentally to anti-correlate with degrees of growth limitation of intracellular metabolites (Reznik et al., 2013), underlining the biological relevance of model parameters relating to the overall flexibility of the system. The same study found that metabolites with negative shadow prices (i.e. that are predicted to be growth-limiting) show lower temporal variation following a perturbation, indicating that *in vivo* their concentrations are more buffered and tightly regulated. These two parameters are related, in that reduced costs of exchange reactions point to uptake constraints on metabolites that are limiting for the solution, while the shadow costs may directly identify the metabolites themselves. In what follows, we focus on the reduced costs of reactions in order to simplify interpretation, since reactions are already annotated with pathways within the model, while metabolites are not directly annotated to pathways.

We calculated the reduced costs for each time point and strain with the model objective and constraints set as per standard FBA, as discussed in the Methods. The overall counts of reactions with reduced costs below zero for each condition are indicated in Supplementary Figure A.10. There is no overall trend with respect to age in any of the strains, however, GLP has a larger number of reactions with reduced cost below zero at most of the time points than the other two strains, providing evidential support indicating that GLP depends on more pathways acting at optimal level than the other strains.

We can further explore the distribution of reactions with reduced costs below zero by performing pathway enrichment to identify pathways that are over-represented for such reactions for each time point and strain, as described in the Methods.

Looking at the pathway enrichments for the reactions with reduced costs below zero, i.e. those reactions that would lead to an increased optimal value if they were able to carry more flux (Figure 5.2), we can see that only a few pathways are regularly enriched for constrained reactions. These pathways are the ones that ‘throttle’ the system as a whole.

A few subtle strain differences are immediately apparent. The GLP strain has enrichments at almost all the time points in mitochondrial transport, while FEM and GEM have few time points with enrichments in this pathway – the older time points in FEM, a few of the younger time points in GEM. In GLP, the fact that the *transport* reactions in particular are throttling the growth of the system can be expected to correspond to the shuttling of

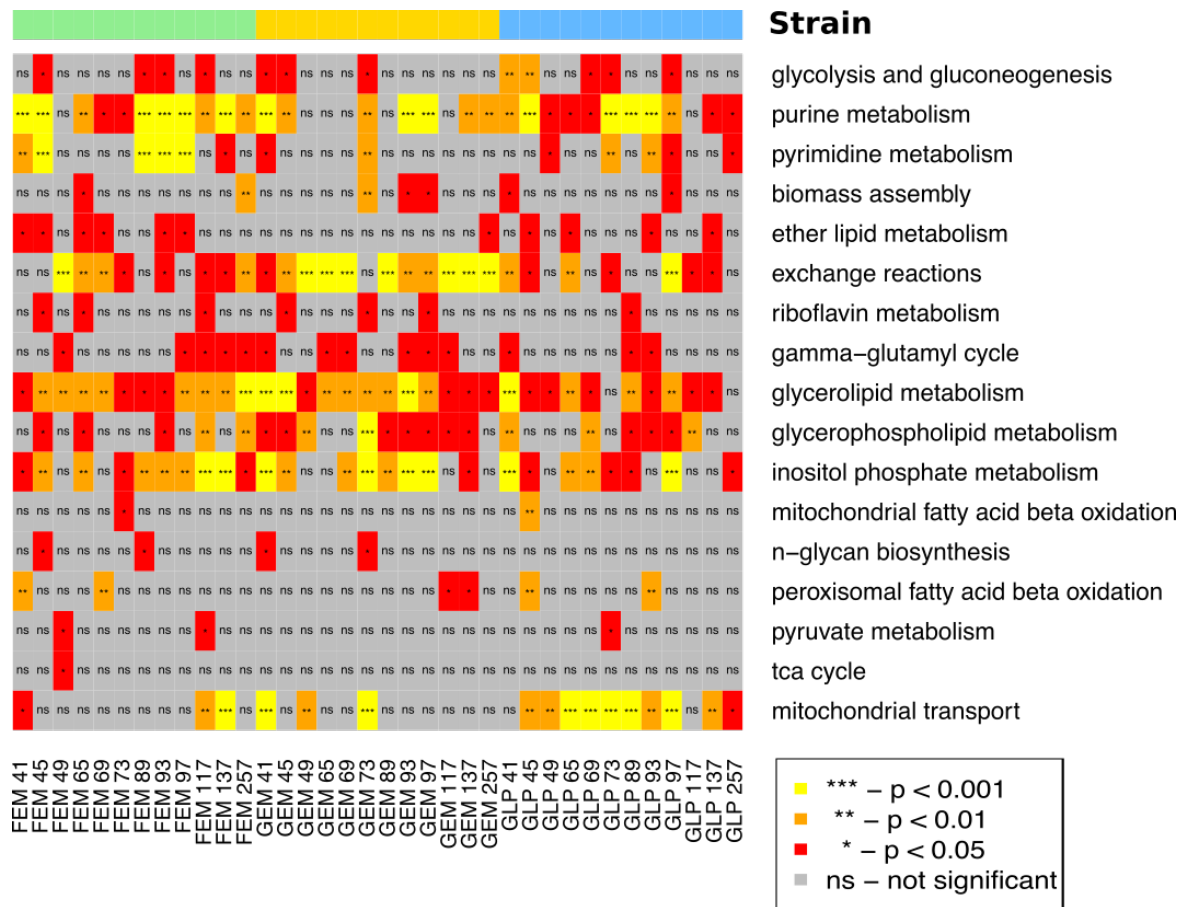


Figure 5.2 Pathway enrichment for reactions with reduced costs below zero

The Figure shows the results of pathway enrichment for the reactions with reduced costs below zero, grouped by strain and then by age. Within the plot, colour indicates significance of enrichment, with yellow the most significant followed by orange, red and then grey as not significant. Column top bar colour indicates strain: green for FEM, gold for GEM and blue for GLP.

byproducts of altered fatty acid metabolism into the mitochondrion for oxidation, although intriguingly mitochondrial fatty acid beta oxidation as a pathway is *not* enriched at most time points in GLP, indicating that it is the transportation rather than the oxidation that is setting the constraint on the system in this case. Exchange reactions as a whole are more enriched, at more time points, in the GEM strain than the FEM and GLP, indicating that this strain might be able to grow more if more inputs were allowed from the environment. Glycolysis and gluconeogenesis seem slightly more enriched in GLP, particularly at the earlier time points in the time series, while purine metabolism is more enriched in FEM during middle age.

During the early ageing time points, we see once again the ‘switching’ behaviour characteristic of the metabolic oscillations that we observed in the metabolome, in this case indicating different distribution of system constraints under the different time of day condi-

tions. For example, riboflavin metabolism, ether lipid metabolism, glycerophospholipid metabolism and inositol phosphate metabolism appear to be behaving that way at least in some of the strains.

Overall, it is apparent that there is considerable complexity in which aspects of the system are more constrained as revealed by this analysis. Almost every time point and strain has a different profile in terms of which combination of pathways are enriched and at what significance, suggesting that different aspects of the overall metabolic flexibility of the organism are more relevant in different conditions and underscoring the importance of detailed, whole-systems approaches to tease apart complex dependencies.

5.3.2 Prediction of novel longevity modulators based on metabolite network distances

As discussed in Chapter 2, several of our metabolites are known longevity modulators – extending lifespan when supplemented. Given the broad and untargeted nature of the measured dataset, without *a priori* selection of metabolites for specific purposes, it may fortuitously allow for the discovery of entirely novel regulators of the ageing process.

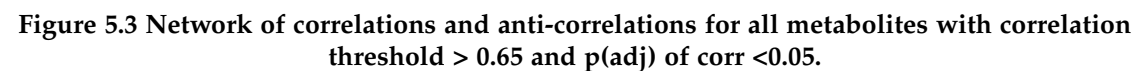
Based on a simple correlation analysis and a threshold, we can induce a network among the metabolites where edges represent correlation and anti-correlation, as illustrated in Figure 5.3. Looking at this simple correlation network, it is clear that some metabolites are more central while others are more peripheral. Moreover, there are potential apparent clusters of longevity modulators, and it can be hypothesised that metabolites that are clustering together with longevity modulators are themselves more likely to be longevity modulators.

To formalise this intuition, we can define a network index d^L for a given metabolite that represents an average of the density of longevity modulators in the immediate neighbourhood of the metabolite, weighted to favour metabolites that are more central and correlations that are more strong, as follows:

$$d^L = 1 \text{ if } d^L \in \{\text{longevity modulators}\} \quad (5.1)$$

$$\text{else } d^L = \frac{\sum_{i=1}^n d^{L_i} \sum_{i=1}^n |c^i|}{(n+1)} \quad (5.2)$$

where n is the number of edges around the metabolite, $i \in \{1, \dots, n\}$ is an index for each edge, and c^i is the correlation strength of that edge. Informally, this measure consists of the sum of the scores as longevity modulators of the immediate network neighbours, modulated by a measure for (local, degree-based) centrality based on the number of edges and the strengths of the correlations. The score for longevity modulators depends on the score for the neighbouring nodes, which in turn depends on the next neighbouring nodes,



but the recursion is only one level deep. Only the strength of correlation matters, not the direction.

Computing this measure for each metabolite in the network gives a score for each metabolite that is 1 for those metabolites that are known longevity modulators, and a value between 0 and 1 for those that are not yet known to be longevity modulators, related to their network neighbourhoods. Higher values indicate higher predictive scores. Metabolites scoring highly on this metric are central and closely related to a number of known longevity modulators, while metabolites scoring lowly are peripheral and not related to any known longevity modulators.

The full table of predictions (for non-longevity-modulating metabolites) is included as Supplementary Table A.7, while the highest and lowest predictions together with the distribution of the score predictions are illustrated in Figure 5.4.

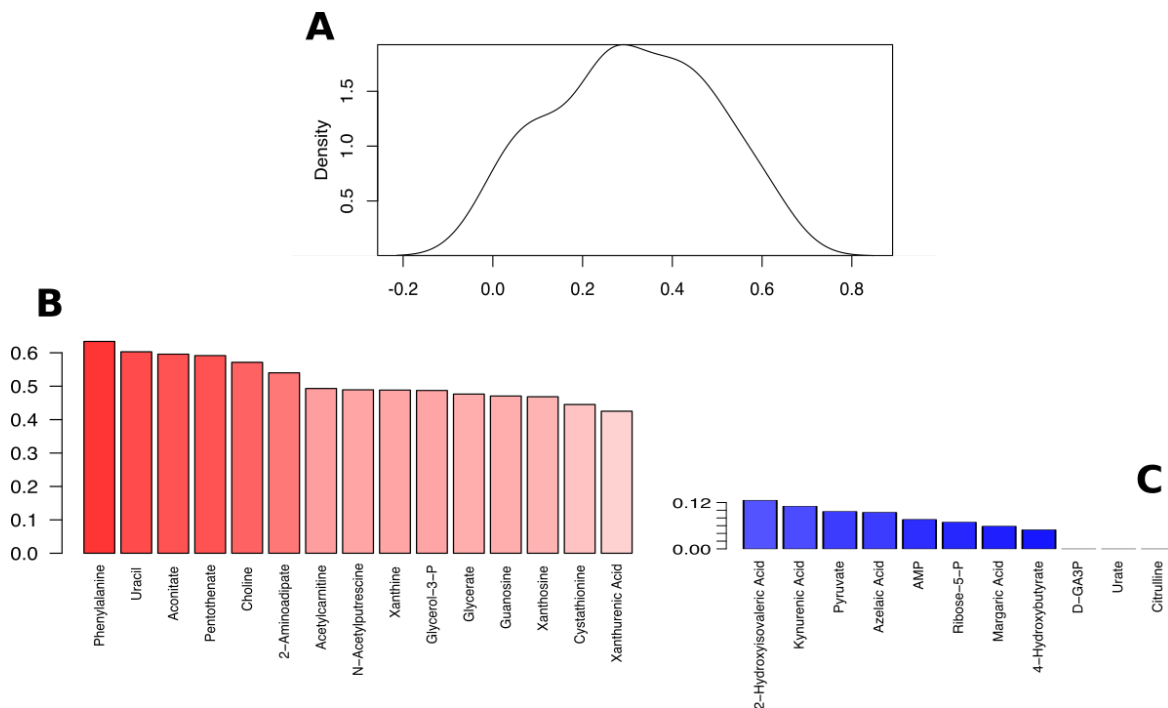


Figure 5.4 Network-based predictions of longevity modulating scores for metabolites

(A) Density of predictions for non-longevity-modulating metabolites, indicating that the metric d^L is fairly evenly distributed in this dataset. (B) The top end of the metric shows the metabolites with the highest predictions. (C) The lower tail end of the metric shows the metabolites with the lowest predictions.

The highest predicted metabolite according to this metric is phenylalanine, an essential amino acid with many known biological roles. At least one study has tested phenylalanine supplementation and *not* found an effect on longevity (Copes et al., 2015), but this finding may not generalise to all strains and conditions.

The next highest are uracil, a nucleobase, aconitate, part of central energy metabolism via the TCA cycle, and pentothenate, a vitamin with complex roles in energy production. Moreover, while this metric did not explicitly take into consideration whether the metabolites were changing significantly with age (Chapter 2), it is of course apparent from the network graph (Figure 5.3) that the highest density nodes are amongst those found to be increasing or decreasing, i.e. the age-associated metabolites, as one might expect.

The top scoring metabolites according to this metric seem eminently plausible candidates for longevity modulation and it is possible to hypothesise that some of the top 10 or 20 might well be novel longevity modulators. On the bottom end of the prediction scale, we

find metabolites that are less likely to be modulators of longevity according to this metric. However, in the absence of experimental testing, how can we best validate if the score is accurate?

One possible validation would be to determine to what extent the metric can predict *known* longevity modulators, if the information that they are known to be such is removed from the dataset (a ‘leave-one-out’ validation design). The results of this experiment are encouraging: for the most part, the known longevity modulators score more highly than the unknown metabolites (Figure 5.5). However, there are some exceptions, and the tail end of the prediction metric validation includes the metabolites spermidine, PEP, cytidine and glycine, which are all non-central in the network and not well connected to other longevity modulators (Figure 5.3). It is interesting to note that while spermidine is a well-known longevity modulator, it was also highly variable in our dataset.

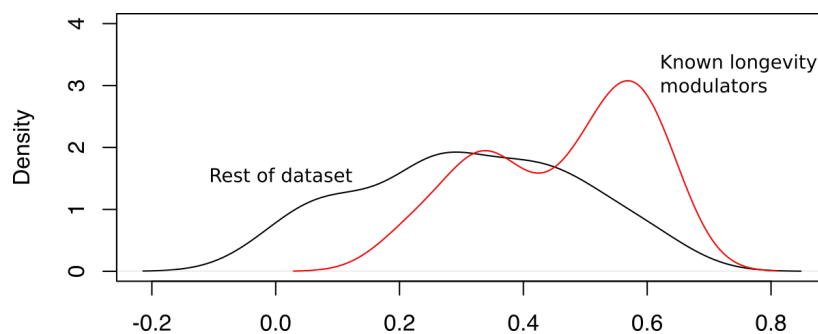


Figure 5.5 Validation of predicted longevity modulating scores for metabolites

The Figure shows the density of predictions for known longevity modulating metabolites, computed in a leave-one-out validation approach, for the longevity modulators (red line) and the remainder of the dataset, i.e. non-longevity-modulators. On average, the known longevity modulators receive a higher predicted longevity modulation score, as would be hoped.

Interestingly, PEP is a known longevity modulator that is not age-associated and in this network analysis is far removed from the main cluster of age-associated metabolites, hinting at different underlying mechanisms. In (Feng et al., 2016; Yuan et al., 2016), a progressive decline in PEP carboxykinase with age and reciprocal increase in pyruvate kinase shunt energy metabolism was observed to lead to a metabolic shift with age from oxidative metabolism to anaerobic glycolysis, a metabolic event that is retarded by caloric restriction.

What more can we learn from considering the metabolites fluctuations together with the gene expression fluctuations from the associated, linked transcriptomics dataset? We evaluate this next.

5.3.3 Exploring mechanisms of metabolic regulation by predicting gene–metabolite interactions

As our study had linked samples from which both transcriptomics and metabolomics data were obtained, it was possible to do a ‘trans-omics’ (Yugi et al., 2016) correlation analysis across the two different omics layers, and thereby induce a simple network with edges linking genes to metabolites.

As there are many more genes in the transcriptomics dataset than metabolites in the metabolomics dataset, this type of analysis may be somewhat uneven in predictive power for the different data types, resulting in a different noise level for genes and metabolites. Therefore, to mitigate this problem, only a subset of genes were used in the analysis: those that have been indicated as having a metabolic role in the analysis of the full time series in our lab (Mains, 2018). The resulting induced network is shown in Figure 5.6.

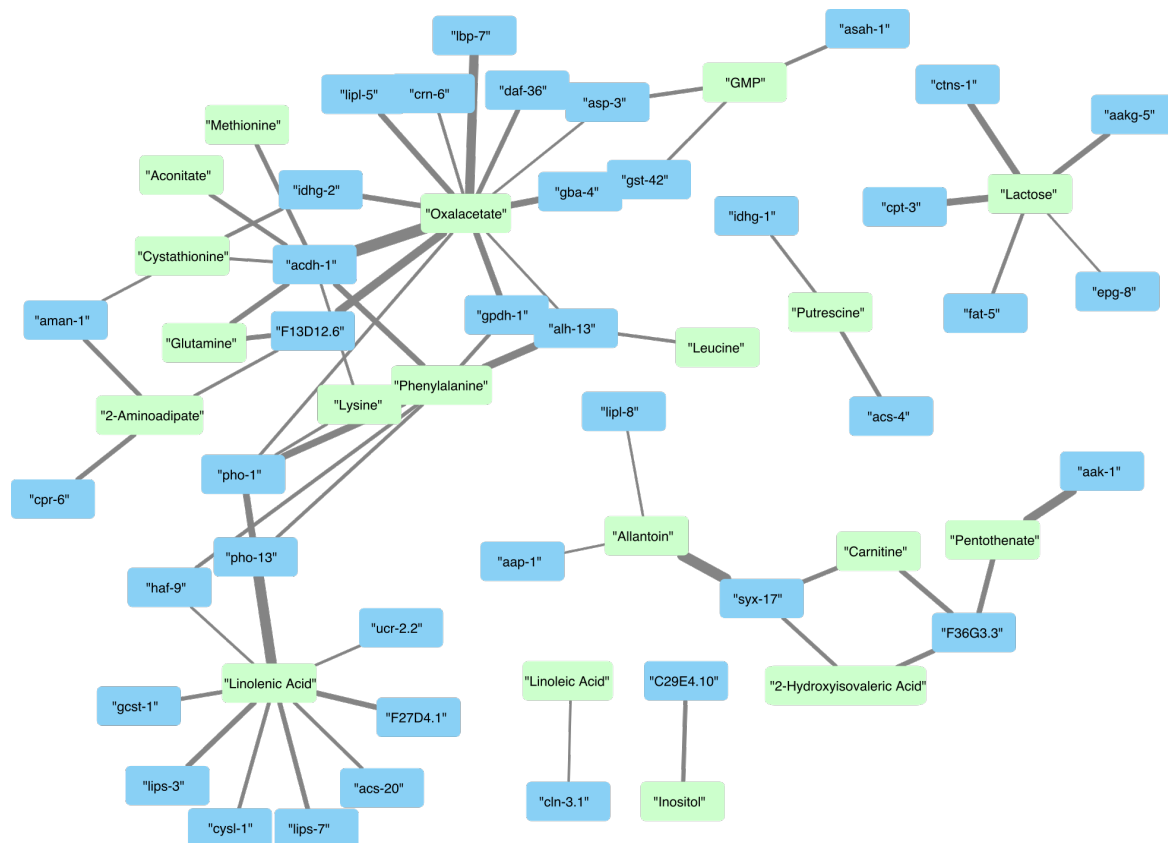


Figure 5.6 Network of correlations between genes and metabolites with correlation threshold > 0.7 and $p(\text{adj})$ of corr < 0.05.

The Figure illustrates the bipartite network of correlations between genes and metabolites with correlation threshold > 0.7 and $p(\text{adj})$ of corr < 0.05. Metabolites are coloured green, while genes are coloured blue. Edge width represents correlation strength.

Based on this network, similarly to the approach we used above for the metabolite–metabolite network, we can define a simple ranking of metabolites for how central they are. Here, the emphasis is on degree centrality, i.e. in terms of how many connections a node has, rather than overall network centrality. The assumption is that the more central the metabolite, the more likely it is to have a potentially regulatory role. The centrality of a metabolite, c^M is defined as:

$$c^M = \sum_{i=1}^n \left(\frac{c^i m}{M} \right) \quad (5.3)$$

where n is the edge count for a particular metabolite, i is an index for that metabolite's edges to genes, c^i is the correlation strength for the particular edge to the gene connected by edge i , m is the count of edges for that correlated gene to further metabolites, and M is the total possible gene-metabolite edges for a single gene across the whole network.

This gives a ranking of the metabolites present in the network (i.e. those that do have above-threshold correlations with genes). The scores are shown in Table 5.1.

| Metabolite | Centrality Score | Longevity modulator |
|--------------------------|------------------|---------------------|
| Oxaloacetate | 3.24 | Yes |
| Phenylalanine | 2.13 | No |
| Linolenic Acid | 1.60 | Yes |
| Cystathionine | 1.12 | No |
| Lysine | 1.11 | Yes |
| Glutamine | 1.05 | Yes |
| Methionine | 0.73 | Yes |
| Aconitate | 0.73 | No |
| Carnitine | 0.63 | Yes |
| 2-Hydroxyisovaleric Acid | 0.62 | No |
| 2-Aminoadipate | 0.62 | No |
| Allantoin | 0.55 | Yes |
| Lactose | 0.53 | No |
| GMP | 0.51 | No |
| Pentothenate | 0.43 | No |
| Leucine | 0.31 | Yes |
| Putrescine | 0.21 | Yes |
| Inositol | 0.10 | Yes |
| Linoleic Acid | 0.10 | No |

Table 5.1 Centrality scores for metabolites in gene–metabolite network.

This is of course just a tentative analysis, as the dataset of metabolites is not necessarily representative of all the changes that are taking place in the metabolism of the organism, and therefore cannot be assumed to reflect an unbiased result. However, it is reassuring to observe that oxaloacetate, which emerged as a key rate-limiting metabolite in central energy

metabolism in our FBA analysis (Chapter 4) and a robust longevity modulator, appears as the highest ranked metabolite for centrality according to this definition in our induced gene–metabolite network.

The genes most closely associated with oxaloacetate in this correlation study include *lbp-7*, an ortholog of human fatty acid binding proteins; *lipl-5*, an intestinal gene with hydrolase activity and a member of a family that also includes *lipl-4*, a known longevity promoter that is required for both germline-free and insulin mutant longevity (Hou and Taubert, 2012); *acdh-1*, an acyl-coA dehydrogenase that catalyses the first step of mitochondrial fatty acid beta-oxidation, and for which reduction of its activity shortens lifespan (RNAi data in WormBase); and *daf-36*, part of the *daf-12* hormonal signalling pathway.

Phenylalanine appears next, in second place, a metabolite that was also highly ranked in the predictive score based on the metabolite-metabolite network above, highlighting that even if this metabolite is not itself a longevity modulator, it must play a key role in the intermediary metabolism that underlies the pathways that lead to longevity. It correlates strongly with *alh-13*, a mitochondrial gene involved in glutamate metabolism, and *pho-1*, an intestinal acid phosphatase, that also correlates with linolenic acid, along with *pho-13*.

Among the further notably strong correlations in this network is pantothenate correlating with *aak-1*, a notable AMP-activated protein kinase that mediates lifespan, oxidative metabolism and fat composition (Greer et al., 2009; Moreno-Arriola et al., 2016; Webster et al., 2017). Pantothenate is not a known longevity modulator, but it is required for normal development. While it has a low score for centrality based on the metabolite–gene correlations, it does obtain a high score as a predicted longevity modulator based on metabolite–metabolite correlations (Figure 5.4B).

5.4 Conclusion

This chapter explored several different network-based and trans-omic approaches to better understand flexibility and regulation within our time series multi-omic dataset. Metabolite–metabolite correlation analysis revealed metabolites that would be good candidates for testing as candidate longevity modulators, while a metabolite–gene correlation analysis revealed trans-omic connections that can be further probed to explore possible cross-layer regulatory connections. The genome-scale modelling approach was used to explore both flux variability under the different conditions of the study (i.e. the different time points in the time series, and different strains) and determine which aspects of the network were the most bottlenecked in the shadow cost analysis. These analyses are exploratory and have not been comprehensively validated, but show some promise for new directions to probe the interrelationships between metabolism and the genetic control of longevity pathways in *C. elegans*.

6. General Discussion

Despite extensive progress in understanding the metabolic and gene regulatory pathways underlying the processes of ageing, its effects remain a pervasive challenge and better treatments an urgent necessity. Systems approaches are needed which are able to link together hugely complex synchronized maps of regulatory connections under different conditions in order to make truly predictive modelling possible in this field (Hastings et al., 2019b; Sweetlove et al., 2014). Technical advances in measurement technologies enable large-scale -omics data to probe many different layers more comprehensively, yet truly integrative analyses still lag behind (Alyass et al., 2015).

In this thesis, I have harnessed several different computational approaches to try to better understand the changes due to ageing in a detailed time series of gene expression and metabolic -omics measurements. I evaluated the changes in metabolome levels due to age and observed a subset of metabolic fluctuations associated with time of day during early ageing. I also compared the metabolome of two normal-lived strains with a long-lived germline-free strain and determined that at least in some aspects, the long-lived strain appeared more youthful for longer, although in other aspects such as differential metabolism of fats, it is quite different. I developed and validated a novel method for the integration of metabolomics data with time series metabolic modelling (Hastings et al., 2019a). This is the first such study to use two different -omics layers integratively together with a genome-scale metabolic network in *C. elegans*, and moreover contributed to the development of a consensus metabolic network that will serve as a resource for the whole community to further this type of systems-based analysis approach (Witting et al., 2018). Using this modelling approach, I was able to predict fluxes for every reaction in the network and compare the flux predictions between different times and strains, and to highlight parts of the network that are more constrained and more flexible.

Several complexities of ageing metabolism have not been covered in my analyses and are thus limitations on this work. I have made no attempt to consider the relationship between *C. elegans* and the metabolism of its bacterial food source, nor the tissue specificity of the various different aspects of metabolism taking place in the organism.

C. elegans are bacteriovores and are usually cultured with *E. coli* as their monoxenic food source. It is known that some of the dietary interventions that affect ageing in *C. elegans* are mediated by active bacterial metabolism, such that the bacterial diet constitutes not only a food source but also an active agent in the phenotypic effects, similarly to the operation of the human microbiome (Cabreiro et al., 2013; Cabreiro and Gems, 2013; Virk et al., 2012, 2016). The provision of different strains of *E. coli* as food source can profoundly affect the worm metabolism and lifespan (MacNeil et al., 2013). It has also been shown that the

pro-longevity effect of the anti-diabetic drug metformin (a calorie restriction mimetic) exerts its effect on the folate and methionine metabolism of the co-culture bacteria (Cabreiro et al., 2013) and a similar pro-longevity effect occurs when *C. elegans* are fed with *E. coli* mutant in the folate pathway (Virk et al., 2012). These findings illustrate that it will be essential to develop composite models which represent both the metabolism of *C. elegans* and its bacterial food source.

In terms of tissue specificity, an important element that is completely overlooked in the approaches presented thus far is the complexity of neuronal control of metabolism. For example, Entchev et al. (2015) discuss an important mechanism for the rewiring of metabolism in response to sensory perception, while Riera and Dillin (2016) discuss further mediators these relationships, altering organismal responses to metabolic signals from the environment.

In addition, several aspects of our approach can be improved, and such improvements remain for future work. Many of the analyses presented in this thesis are exploratory and need to be further validated on other datasets, as well as – crucially – tested experimentally. The set of metabolites available for the study was limited and may have introduced bias to the results. In future a more comprehensive, untargeted metabolomics assay might be used to obtain a whole-metabolome view on systemic changes due to ageing.

Another significant limitation of our study is the incomplete state of annotation of the model of *C. elegans* metabolism. While the WormJam model represents a consensus of the knowledge represented across all the available published models, it is a work in progress and there are known problems which require further manual curation to resolve (Witting et al., 2018). For example, the annotation of worm-specific metabolites is poor, as is the annotation of pathways involving fatty acids and lipids. Large-scale community involvement is essential for this to succeed, as no single group has access to all the expertise that is required for such a large scale knowledge building activity to be brought up to date with everything that is currently known in the field. Moreover, visualisation of predicted fluxes in such a large-scale network remains a challenge, and although tools such as Escher do exist, this is an area which urgently needs more sophisticated tooling.

Crucially, we anticipate that other approaches to changing the objective function will also enhance the use of FBA for the study of ageing in *C. elegans* (Hastings et al., 2019b). For example, the addition of *in vivo* measured oxygen consumption and total ATP production data at matching timepoints would enable the accurate scaling of central energy metabolism to better match the physiological decline that is known to occur with age (Braeckman et al., 2002), yet is not fully transcriptionally evident. This would force the model to find other pathways for energy generation and would no doubt lead to further insights into the network-based complexity of the loss of functioning during the ageing process.

Systems biology and multi-omics approaches to the study of ageing in model organisms such as *C. elegans* have the potential to reveal novel insights about mammalian and human ageing processes, as most of the genes and pathways involved in ageing worms are conserved and have similar effects in humans. Age-associated phenotypes such as the accumulation of molecular and cellular damage, the decline of physiological functions such as energy generation and stress resistance, and the increase in risk of disease morbidity and mortality, are also shared between model organisms and humans, driven by the same conserved genes. However, model organisms are experimentally tractable and are thus able to be used in large-scale screening studies that lead to the development of novel interventions targeting the extension of health and longevity in humans. Both for basic discovery of novel aspects of the biology of ageing, and for the more applied challenge in determining novel treatments, good *in silico* models that are able to represent the physiological environment and make predictions based on large-scale omics measurements in model organisms are important, and these predictions can be translated to humans via gene orthology.

Many of the findings of this study recapitulate ageing-associated effects that are known to occur similarly in humans. For example, the depletion of amino acids and decline of TCA cycle function that we observe with age in *C. elegans* is paralleled by a decline in metabolic function generally and in particular in the maintenance of muscle tissue, which forms the main amino acid repository of the human body (Timmerman and Volpi, 2010). Using systems biology approaches to tease apart the relative contribution of each of these different aspects to systemic phenotypes enables specific causal predictions to be made for the different interacting pathways. One of the main contributions of this thesis is to enhance the systems biology predictive capability with a more comprehensive model of *C. elegans* metabolism during ageing. While there are limits to what can be inferred from *C. elegans* to humans, as some aspects of *C. elegans* ageing physiology – such as the excess of yolk production post-reproduction – are not conserved (Gruber et al., 2015), it nevertheless provides an efficient and tractable system for manipulation and discovery.

References

- Ackerman, D. and Gems, D. (2012). The mystery of *C. elegans* aging: an emerging role for fat. Distant parallels between *C. elegans* aging and metabolic syndrome? *BioEssays*, 34:466–471.
- Aguilaniu, H. (2015). The mysterious relationship between reproduction and longevity. *Worm*, 4:e1020276.
- Akesson, M., Förster, J., and Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. *Metabolic Engineering*, 6(4):285–293.
- Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z. N., and Barabási, A.-L. (2004). Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, 427(6977):839–843.
- Alvares, S. M., Mayberry, G. A., Joyner, E. Y., Lakowski, B., and Ahmed, S. (2014). H3K4 demethylase activities repress proliferative and postmitotic aging. *Aging Cell*, 13:245–253.
- Alyass, A., Turcotte, M., and Meyre, D. (2015). From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Medical Genomics*, 8:33.
- Amrit, F. R. G., Steenkiste, E. M., Ratnappan, R., Chen, S.-W., McClendon, T. B., Kostka, D., Yanowitz, J., Olsen, C. P., and Ghazi, A. (2016). DAF-16 and TCER-1 facilitate adaptation to germline loss by restoring lipid homeostasis and repressing reproductive physiology in *C. elegans*. *PLoS Genetics*, 12:e1005788.
- Anderson, E. N., Corkins, M. E., Li, J.-C., Singh, K., Parsons, S., Tucey, T. M., Sorkaç, A., Huang, H., Dimitriadi, M., Sinclair, D. A., and Hart, A. C. (2016). *C. elegans* lifespan extension by osmotic stress requires FUDR, base excision repair, FOXO, and sirtuins. *Mechanisms of Ageing and Development*, 154:30–42.
- Apfeld, J., O'Connor, G., McDonagh, T., DiStefano, P. S., and Curtis, R. (2004). The AMP-activated protein kinase AAK-2 links energy levels and insulin-like signals to lifespan in *C. elegans*. *Genes & Development*, 18:3004–3009.
- Armitage, E. G., Godzien, J., Alonso-Herranz, V., López-González, A., and Barbas, C. (2015). Missing value imputation strategies for metabolomics data. *Electrophoresis*, 36:3050–3060.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. the Gene Ontology Consortium. *Nature Genetics*, 25:25–29.
- Back, P., De Vos, W. H., Depuydt, G. G., Matthijssens, F., Vanfleteren, J. R., and Braeckman, B. P. (2012). Exploring real-time in vivo redox biology of developing and aging *Caenorhabditis elegans*. *Free Radical Biology & Medicine*, 52:850–859.
- Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L.-S. L. (2005). The Universal Protein Resource (UniProt). *Nucleic Acids Research*, 33(Database issue):D154–D159.
- Barardo, D., Thornton, D., Thoppil, H., Walsh, M., Sharifi, S., Ferreira, S., Anžič, A., Fernandes, M., Monteiro, P., Grum, T., Cordeiro, R., De-Souza, E. A., Budovsky, A., Araujo, N., Gruber, J., Petrascheck, M., Fraifeld, V. E., Zhavoronkov, A., Moskalev, A., and de Magalhães, J. P. (2017a). The DrugAge database of aging-related drugs. *Aging Cell*, 16:594–597.
- Barardo, D. G., Newby, D., Thornton, D., Ghafourian, T., de Magalhães, J. P., and Freitas, A. A. (2017b). Machine learning for predicting lifespan-extending chemical compounds. *Aging*, 9:1721–1737.
- Bartel, J., Krumsiek, J., Schramm, K., Adamski, J., Gieger, C., Herder, C., Carstensen, M., Peters, A., Rathmann, W., Roden, M., Strauch, K., Suhre, K., Kastenmüller, G., Prokisch, H., and Theis, F. J. (2015). The human blood metabolome-transcriptome interface. *PLoS Genetics*, 11(6):e1005274.
- Bartel, J., Krumsiek, J., and Theis, F. J. (2013). Statistical methods for the analysis of high-throughput metabolomics data. *Computational and Structural Biotechnology*, 4:e201301009.
- Barton, M. K., Schedl, T. B., and Kimble, J. (1987). Gain-of-function mutations of *fem-3*, a sex-determination gene in *Caenorhabditis elegans*. *Genetics*, 115:107–119.

- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.
- Bathi, R. J., Parveen, S., Mutalik, S., and Rao, R. (2010). Rabson-Mendenhall syndrome: two case reports and a brief review of the literature. *Odontology*, 98:89–96.
- Beard, J. R. and Bloom, D. E. (2015). Towards a comprehensive public health response to population ageing. *Lancet*, 385:658–661.
- Becker, S. A. and Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology*, 4(5):e1000082.
- Beger, R. D., Dunn, W., Schmidt, M. A., Gross, S. S., Kirwan, J. A., Cascante, M., Brennan, L., Wishart, D. S., Oresic, M., Hankemeier, T., Broadhurst, D. I., Lane, A. N., Suhre, K., Kastenmüller, G., Sumner, S. J., Thiele, I., Fiehn, O., Kaddurah-Daouk, R., for "Precision Medicine, and Initiative, P. T. G.-M. S. (2016). Metabolomics enables precision medicine: "a white paper, community perspective". *Metabolomics*, 12:149.
- Beilharz, J. E., Maniam, J., and Morris, M. J. (2015). Diet-induced cognitive deficits: The role of fat and sugar, potential mechanisms and nutritional interventions. *Nutrients*, 7(8):6719–6738.
- Ben-Zvi, A., Miller, E. A., and Morimoto, R. I. (2009). Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *Proceedings of the National Academy of Sciences of the United States of America*, 106:14914–14919.
- Berman, J. R. and Kenyon, C. (2006). Germ-cell loss extends *C. elegans* life span through regulation of DAF-16 by kri-1 and lipophilic-hormone signaling. *Cell*, 124:1055–1068.
- Bjelakovic, G., Nikolova, D., Gluud, L. L., Simonetti, R. G., and Gluud, C. (2007). Mortality in randomized trials of antioxidant supplements for primary and secondary prevention: systematic review and meta-analysis. *Journal of the American Medical Association*, 297:842–857.
- Blackwell, T. K., Steinbaugh, M. J., Hourihan, J. M., Ewald, C. Y., and Isik, M. (2015). SKN-1/Nrf, stress responses, and aging in *Caenorhabditis elegans*. *Free Radical Biology & Medicine*, 88:290–301.
- Bordbar, A., Yurkovich, J. T., Paglia, G., Rolfsson, O., Sigurjónsson, O. E., and Palsson, B. O. (2017). Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics. *Scientific Reports*, 7:46249.
- Braeckman, B. P., Houthoofd, K., De Vreese, A., and Vanfleteren, J. R. (2002). Assaying metabolic activity in ageing *Caenorhabditis elegans*. *Mechanisms of Ageing and Development*, 123:105–119.
- Braeckman, B. P., Houthoofd, K., and Vanfleteren, J. R. (2009). Intermediary metabolism. *WormBook*.
- Braeckman, B. P. and Vanfleteren, J. R. (2007). Genetic control of longevity in *C. elegans*. *Experimental Gerontology*, 42:90–98.
- Brereton, R. G. and Lloyd, G. R. (2014). Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics*, 28:213–225.
- Brereton, R. G. and Lloyd, G. R. (2018). Partial least squares discriminant analysis for chemometrics and metabolomics: How scores, loadings, and weights differ according to two common algorithms. *Journal of Chemometrics*, 32:e3028.
- Brown, A., Guess, N., Dornhorst, A., Taheri, S., and Frost, G. (2017). Insulin-associated weight gain in obese type 2 diabetes mellitus patients: What can be done? *Diabetes, Obesity & Metabolism*, 19:1655–1668.
- Brunquell, J., Bowers, P., and Westerheide, S. D. (2014). Fluorodeoxyuridine enhances the heat shock response and decreases polyglutamine aggregation in an HSF-1-dependent manner in *Caenorhabditis elegans*. *Mechanisms of Ageing and Development*, 141-142:1–4.
- Brys, K., Castelein, N., Matthijssens, F., Vanfleteren, J. R., and Braeckman, B. P. (2010). Disruption of insulin signalling preserves bioenergetic competence of mitochondria in ageing *Caenorhabditis elegans*. *BMC Biology*, 8:91.
- Burgard, A. P., Vaidyaraman, S., and Maranas, C. D. (2001). Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnology Progress*, 17(5):791–797.
- Burnett, C., Valentini, S., Cabreiro, F., Goss, M., Somogyvári, M., Piper, M. D., Hoddinott, M., Sutphin, G. L., Leko, V., McElwee, J. J., Vazquez-Manrique, R. P., Orfila, A.-M., Ackerman, D., Au, C., Vinti, G., Riesen, M., Howard, K., Neri, C., Bedalov, A., Kaeberlein, M., Soti, C., Partridge, L., and Gems, D. (2011). Absence of effects of Sir2 overexpression on lifespan in *C. elegans* and *Drosophila*. *Nature*, 477:482–485.
- Bárcena, C., Mayoral, P., and Quirós, P. M. (2018). Mitohormesis, an antiaging paradigm. *International Review of Cell and Molecular Biology*, 340:35–77.
- Cabreiro, F., Au, C., Leung, K.-Y., Vergara-Irigaray, N., Cochemé, H. M., Noori, T., Weinkove, D., Schuster, E., Greene, N. D. E., and Gems, D. (2013). Metformin retards aging in *C. elegans* by altering microbial folate and methionine metabolism. *Cell*, 153(1):228–239.

- Cabreiro, F. and Gems, D. (2013). Worms need microbes too: microbiota, health and aging in *Caenorhabditis elegans*. *EMBO Molecular Medicine*, 5(9):1300–1310.
- Caito, S. W. and Aschner, M. (2015). Quantification of glutathione in *Caenorhabditis elegans*. *Current Protocols in Toxicology*, 64:6.18.1–6.18.6.
- Calvert, S., Tacutu, R., Sharifi, S., Teixeira, R., Ghosh, P., and de Magalhães, J. P. (2016). A network pharmacology approach reveals new candidate caloric restriction mimetics in *C. elegans*. *Aging Cell*, 15:256–266.
- Carling, D. (2017). AMPK signalling in health and disease. *Current Opinion in Cell Biology*, 45:31–37.
- Carneiro, G., Radcenco, A. L., Evaristo, J., and Monnerat, G. (2019). Novel strategies for clinical investigation and biomarker discovery: a guide to applied metabolomics. *Hormone Molecular Biology and Clinical Investigation*.
- Casella, R., Evangelisti, E., Zampagni, M., Becatti, M., D’Adamio, G., Goti, A., Liguri, G., Fiorillo, C., and Cecchi, C. (2014). S-linolenoyl glutathione intake extends life-span and stress resistance via Sir-2.1 upregulation in *Caenorhabditis elegans*. *Free Radical Biology & Medicine*, 73:127–135.
- Chaudhari, S. N. and Kipreos, E. T. (2017). Increased mitochondrial fusion allows the survival of older animals in diverse *C. elegans* longevity pathways. *Nature Communications*, 8:182.
- Chaudhari, S. N. and Kipreos, E. T. (2018). The energy maintenance theory of aging: Maintaining energy metabolism to allow longevity. *BioEssays*, 40:e1800005.
- Chauhan, A., Liebal, U. W., Vera, J., Baltrusch, S., Junghans, C., Tiedge, M., Fuellen, G., Wolkenhauer, O., and Köhling, R. (2015). Systems biology approaches in aging research. *Interdisciplinary Topics in Gerontology*, 40:155–176.
- Chedraui, P. and Pérez-López, F. R. (2013). Nutrition and health during mid-life: searching for solutions and meeting challenges for the aging population. *Climacteric*, 16 Suppl 1:85–95.
- Chen, N., Harris, T. W., Antoshechkin, I., Bastiani, C., Bieri, T., Blasiar, D., Bradnam, K., Canaran, P., Chan, J., Chen, C.-K., Chen, W. J., Cunningham, F., Davis, P., Kenny, E., Kishore, R., Lawson, D., Lee, R., Muller, H.-M., Nakamura, C., Pai, S., Ozersky, P., Petcherski, A., Rogers, A., Sabo, A., Schwarz, E. M., Van Auken, K., Wang, Q., Durbin, R., Spieth, J., Sternberg, P. W., and Stein, L. D. (2005). WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Research*, 33:D383–D389.
- Chin, R. M., Fu, X., Pai, M. Y., Vergnes, L., Hwang, H., Deng, G., Diep, S., Lomenick, B., Meli, V. S., Monsalve, G. C., Hu, E., Whelan, S. A., Wang, J. X., Jung, G., Solis, G. M., Fazlollahi, F., Kaweeteerawat, C., Quach, A., Nili, M., Krall, A. S., Godwin, H. A., Chang, H. R., Faull, K. F., Guo, F., Jiang, M., Trauger, S. A., Saghatelian, A., Braas, D., Christofk, H. R., Clarke, C. F., Teitell, M. A., Petrascheck, M., Reue, K., Jung, M. E., Frand, A. R., and Huang, J. (2014). The metabolite alpha-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. *Nature*, 510:397–401.
- Chung, J. H., Manganiello, V., and Dyck, J. R. B. (2012). Resveratrol as a calorie restriction mimetic: therapeutic implications. *Trends in Cell Biology*, 22(10):546–554.
- Cohen, A. A. (2016). Complex systems dynamics in aging: new evidence, continuing questions. *Biogerontology*, 17:205–220.
- Colijn, C., Brandes, A., Zucker, J., Lun, D. S., Weiner, B., Farhat, M. R., Cheng, T.-Y., Moody, D. B., Murray, M., and Galagan, J. E. (2009). Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Computational Biology*, 5:e1000489.
- Collins, J. J., Huang, C., Hughes, S., and Kornfeld, K. (2008). The measurement and analysis of age-related changes in *Caenorhabditis elegans*. *WormBook*, pages 1–21.
- Colman, R. J., Anderson, R. M., Johnson, S. C., Kastman, E. K., Kosmatka, K. J., Beasley, T. M., Allison, D. B., Cruzen, C., Simmons, H. A., Kemnitz, J. W., and Weindruch, R. (2009). Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science*, 325(5937):201–204.
- Copes, N., Edwards, C., Chaput, D., Saifee, M., Barjuca, I., Nelson, D., Paraggio, A., Saad, P., Lipps, D., Stevens, S. M., and Bradshaw, P. C. (2015). Metabolome and proteome changes with aging in *Caenorhabditis elegans*. *Experimental Gerontology*, 72:67–84.
- Corsi, A. K., Wightman, B., and Chalfie, M. (2015). A transparent window into biology: A primer on *Caenorhabditis elegans*. *Genetics*, 200:387–407.
- Cottret, L., Frainay, C., Chazalviel, M., Cabanettes, F., Gloaguen, Y., Camenen, E., Merlet, B., Heux, S., Portais, J.-C., Poupin, N., Vinson, F., and Jourdan, F. (2018). MetExplore: collaborative edition and exploration of metabolic networks. *Nucleic Acids Research*, 46:W495–W502.
- da Silveira, T. L., Zamberlan, D. C., Arantes, L. P., Machado, M. L., da Silva, T. C., Câmara, D. d. F., Santamaría, A., Aschner, M., and Soares, F. A. A. (2018). Quinolinic acid and glutamatergic neurodegeneration in *Caenorhabditis elegans*. *Neurotoxicology*, 67:94–101.
- Dabbish, N. S. and Raizen, D. M. (2011). GABAergic synaptic plasticity during a developmentally regulated sleep-like state in *C. elegans*. *The Journal of Neuroscience*, 31:15932–15943.

- Dato, S., Crocco, P., D'Aquila, P., de Rango, F., Bellizzi, D., Rose, G., and Passarino, G. (2013). Exploring the role of genetic variability and lifestyle in oxidative stress response for healthy aging and longevity. *International Journal of Molecular Sciences*, 14:16443–16472.
- Davies, S. K., Bundy, J. G., and Leroi, A. M. (2015). Metabolic youth in middle age: Predicting aging in *Caenorhabditis elegans* using metabolomics. *Journal of Proteome Research*, 14:4603–4609.
- Davies, S. K., Leroi, A. M., and Bundy, J. G. (2012). Fluorodeoxyuridine affects the identification of metabolic responses to daf-2 status in *Caenorhabditis elegans*. *Mechanisms of Ageing and Development*, 133:46–49.
- Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M., and Ashburner, M. (2008). ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 36(Database issue):D344–D350.
- Denzel, M. S., Lapierre, L. R., and Mack, H. I. D. (2019). Emerging topics in *C. elegans* aging research: Transcriptional regulation, stress response and epigenetics. *Mechanisms of Ageing and Development*, 177:4–21.
- Di Guida, R., Engel, J., Allwood, J. W., Weber, R. J. M., Jones, M. R., Sommer, U., Viant, M. R., and Dunn, W. B. (2016). Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics*, 12:93.
- Dial, A. G., Ng, S. Y., Manta, A., and Ljubicic, V. (2018). The role of AMPK in neuromuscular biology and disease. *Trends in Endocrinology and Metabolism*, 29:300–312.
- Dikicioglu, D., Kırdar, B., and Oliver, S. G. (2015). Biomass composition: the "elephant in the room" of metabolic modelling. *Metabolomics*, 11(6):1690–1701.
- Dyar, K. A., Lutter, D., Artati, A., Ceglia, N. J., Liu, Y., Armenta, D., Jastroch, M., Schneider, S., de Mateo, S., Cervantes, M., Abbondante, S., Tognini, P., Orozco-Solis, R., Kinouchi, K., Wang, C., Swerdloff, R., Nadeef, S., Masri, S., Magistretti, P., Orlando, V., Borrelli, E., Uhlenhaut, N. H., Baldi, P., Adamski, J., Tschöp, M. H., Eckel-Mahan, K., and Sassone-Corsi, P. (2018). Atlas of circadian metabolism reveals system-wide coordination and communication between clocks. *Cell*, 174:1571–1585.e11.
- Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7:74.
- Edwards, C., Canfield, J., Copes, N., Brito, A., Rehan, M., Lipps, D., Brunquell, J., Westerheide, S. D., and Bradshaw, P. C. (2015). Mechanisms of amino acid-mediated lifespan extension in *Caenorhabditis elegans*. *BMC Genetics*, 16:8.
- Edwards, C. B., Copes, N., Brito, A. G., Canfield, J., and Bradshaw, P. C. (2013). Malate and fumarate extend lifespan in *Caenorhabditis elegans*. *PLoS One*, 8:e58345.
- Eisenberg, T., Knauer, H., Schauer, A., Büttner, S., Ruckenstein, C., Carmona-Gutierrez, D., Ring, J., Schroeder, S., Magnes, C., Antonacci, L., Fussi, H., Deszcz, L., Hartl, R., Schraml, E., Criollo, A., Megalou, E., Weiskopf, D., Laun, P., Heeren, G., Breitenbach, M., Grubeck-Loebenstien, B., Herker, E., Fahrenkrog, B., Fröhlich, K.-U., Sinner, F., Tavernarakis, N., Minois, N., Kroemer, G., and Madeo, F. (2009). Induction of autophagy by spermidine promotes longevity. *Nature Cell Biology*, 11:1305–1314.
- Entchev, E. V., Patel, D. S., Zhan, M., Steele, A. J., Lu, H., and Ch'ng, Q. (2015). A gene-expression-based neural code for food abundance that modulates lifespan. *Elife*, 4:e06259.
- Evason, K., Collins, J. J., Huang, C., Hughes, S., and Kornfeld, K. (2008). Valproic acid extends *Caenorhabditis elegans* lifespan. *Aging Cell*, 7(3):305–317.
- Ezcurra, M., Benedetto, A., Sornda, T., Gilliat, A. F., Au, C., Zhang, Q., van Schelt, S., Petrache, A. L., Wang, H., de la Guardia, Y., Bar-Nun, S., Tyler, E., Wakelam, M. J., and Gems, D. (2018). *C. elegans* eats its own intestine to make yolk leading to multiple senescent pathologies. *Current Biology*, 28:2544–2556.e5.
- Fabris, F., Magalhães, J. P. d., and Freitas, A. A. (2017). A review of supervised machine learning applied to ageing research. *Biogerontology*, 18:171–188.
- Feldman, N., Kosolapov, L., and Ben-Zvi, A. (2014). Fluorodeoxyuridine improves *Caenorhabditis elegans* proteostasis independent of reproduction onset. *PLoS One*, 9:e85964.
- Feng, Z., Hanson, R. W., Berger, N. A., and Trubitsyn, A. (2016). Reprogramming of energy metabolism as a driver of aging. *Oncotarget*, 7(13):15410–15420.
- Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics*, 2:155–168.
- Fiehn, O. (2002). Metabolomics—the link between genotypes and phenotypes. *Plant Molecular Biology*, 48:155–171.
- Finkel, T. (2015). The metabolic regulation of aging. *Nature Medicine*, 21(12):1416–1423.
- Fontana, L., Partridge, L., and Longo, V. D. (2010). Extending healthy life span—from yeast to humans. *Science*, 328(5976):321–326.
- Friedman, D. B. and Johnson, T. E. (1988). A mutation in the age-1 gene in *Caenorhabditis elegans* lengthens life and reduces hermaphrodite fertility. *Genetics*, 118:75–86.

- Fuchs, S., Bundy, J. G., Davies, S. K., Viney, J. M., Swire, J. S., and Leroi, A. M. (2010). A metabolic signature of long life in *Caenorhabditis elegans*. *BMC Biology*, 8:14.
- Gao, A. W., Chatzispyrou, I. A., Kamble, R., Liu, Y. J., Herzog, K., Smith, R. L., van Lenthe, H., Vervaart, M. A. T., van Cruchten, A., Luyf, A. C., van Kampen, A., Pras-Raves, M. L., Vaz, F. M., and Houtkooper, R. H. (2017). A sensitive mass spectrometry platform identifies metabolic changes of life history traits in *C. elegans*. *Scientific Reports*, 7:2408.
- Gao, A. W., Smith, R. L., van Weeghel, M., Kamble, R., Janssens, G. E., and Houtkooper, R. H. (2018). Identification of key pathways and metabolic fingerprints of longevity in *C. elegans*. *Experimental Gerontology*, 113:128–140.
- Garcia, D. and Shaw, R. J. (2017). AMPK: Mechanisms of cellular energy sensing and restoration of metabolic balance. *Molecular Cell*, 66:789–800.
- Gebauer, J., Gentsch, C., Mansfeld, J., Schmeißer, K., Waschina, S., Brandes, S., Klimmasch, L., Zamboni, N., Zarse, K., Schuster, S., Ristow, M., Schäuble, S., and Kaleta, C. (2016). A genome-scale database and reconstruction of *Caenorhabditis elegans* metabolism. *Cell Systems*, 2(5):312–322.
- Gerisch, B., Rottiers, V., Li, D., Motola, D. L., Cummins, C. L., Lehrach, H., Mangelsdorf, D. J., and Antebi, A. (2007). A bile acid-like steroid modulates *Caenorhabditis elegans* lifespan through nuclear receptor signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 104:5014–5019.
- Gerisch, B., Weitzel, C., Kober-Eisermann, C., Rottiers, V., and Antebi, A. (2001). A hormonal signaling pathway influencing *C. elegans* metabolism, reproductive development, and life span. *Developmental Cell*, 1:841–851.
- Glenn, C. F., Chow, D. K., David, L., Cooke, C. A., Gami, M. S., Iser, W. B., Hanselman, K. B., Goldberg, I. G., and Wolkow, C. A. (2004). Behavioral deficits during early stages of aging in *Caenorhabditis elegans* result from locomotory deficits possibly linked to muscle frailty. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59:1251–1260.
- Goya, M. E., Romanowski, A., Caldart, C. S., Bénard, C. Y., and Golombek, D. A. (2016). Circadian rhythms identified in *Caenorhabditis elegans* by in vivo long-term monitoring of a bioluminescent reporter. *Proceedings of the National Academy of Sciences of the United States of America*, 113:E7837–E7845.
- Greer, E. L., Banko, M. R., and Brunet, A. (2009). AMP-activated protein kinase and FoxO transcription factors in dietary restriction-induced longevity. *Annals of the New York Academy of Sciences*, 1170:688–692.
- Greer, E. L., Becker, B., Latza, C., Antebi, A., and Shi, Y. (2016). Mutation of *C. elegans* demethylase spr-5 extends transgenerational longevity. *Cell Research*, 26:229–238.
- Greer, E. L. and Brunet, A. (2009). Different dietary restriction regimens extend lifespan by both independent and overlapping genetic pathways in *C. elegans*. *Aging Cell*, 8:113–127.
- Greer, E. L., Maures, T. J., Ucar, D., Hauswirth, A. G., Mancini, E., Lim, J. P., Benayoun, B. A., Shi, Y., and Brunet, A. (2011). Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature*, 479:365–371.
- Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., and Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta*, 879:10–23.
- Gromski, P. S., Xu, Y., Kotze, H. L., Correa, E., Ellis, D. I., Armitage, E. G., Turner, M. L., and Goodacre, R. (2014). Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites*, 4:433–452.
- Gruber, J., Chen, C.-B., Fong, S., Ng, L. F., Teo, E., and Halliwell, B. (2015). *Caenorhabditis elegans*: What we can and cannot learn from aging worms. *Antioxidants & Redox Signaling*, 23:256–279.
- Gut, P. and Verdin, E. (2013). The nexus of chromatin regulation and intermediary metabolism. *Nature*, 502(7472):489–498.
- Hallen, A., Jamie, J. F., and Cooper, A. J. L. (2013). Lysine metabolism in mammalian brain: an update on the importance of recent discoveries. *Amino Acids*, 45:1249–1272.
- Han, S., Schroeder, E. A., Silva-García, C. G., Hebestreit, K., Mair, W. B., and Brunet, A. (2017). Mono-unsaturated fatty acids link H3K4me3 modifiers to *C. elegans* lifespan. *Nature*, 544:185–190.
- Hansen, M. and Kennedy, B. K. (2016). Does longer lifespan mean longer healthspan? *Trends in Cell Biology*, 26:565–568.
- Harris, J. E., Sheean, P. M., Gleason, P. M., Bruemmer, B., and Boushey, C. (2012). Publishing nutrition research: a review of multivariate techniques—part 2: analysis of variance. *Journal of the Academy of Nutrition and Dietetics*, 112:90–98.
- Hastings, J., Mains, A., Artal-Sanz, M., Bergmann, S., Braeckman, B. P., Bundy, J., Cabreiro, F., Dobson, P., Ebert, P., Hattwell, J., Hefzi, H., Houtkooper, R. H., Jelier, R., Joshi, C., Kothamachu, V. B., Lewis, N., Lourenço, A. B., Nie, Y., Norvaisas, P., Pearce, J., Riccio, C., Rodriguez, N., Santermans, T., Scarcia, P., Schirra, H. J.,

- Sheng, M., Smith, R., Suriyalaksh, M., Towbin, B., Tuli, M. A., van Weeghel, M., Weinkove, D., Zečić, A., Zimmermann, J., le Novère, N., Kaleta, C., Witting, M., and Casanueva, O. (2017). WormJam: A consensus *C. elegans* metabolic reconstruction and metabolomics community and workshop series. *Worm*, 6:e1373939.
- Hastings, J., Mains, A., Virk, B., Rodriguez, N., Murdoch, S., Pearce, J., Bergmann, S., le Novère, N., and Casanueva, O. (2019a). Multi-omics and genome-scale modeling reveal a metabolic shift during *C. elegans* ageing. *Frontiers in Molecular Biosciences*, 6:2.
- Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P., and Steinbeck, C. (2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Research*, 44:D1214–D1219.
- Hastings, J., Suriyalaksh, M., and Casanueva, O. (2019b). Flow with the flux: systems biology tools predict metabolic drivers of ageing in *C. elegans*. *Current Opinion in Systems Biology*, 13:102–107.
- Hendriks, G.-J., Gaidatzis, D., Aeschmann, F., and Großhans, H. (2014). Extensive oscillatory gene expression during *C. elegans* larval development. *Molecular Cell*, 53:380–392.
- Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M. L., Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D. S., Westerhoff, H. V., Kirdar, B., Penttilä, M., Klipp, E., Palsson, B. Ø., Sauer, U., Oliver, S. G., Mendes, P., Nielsen, J., and Kell, D. B. (2008). A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology*, 26(10):1155–1160.
- Hill, S. M., Heiser, L. M., Cokelaer, T., Unger, M., Nesser, N. K., Carlin, D. E., Zhang, Y., Sokolov, A., Paull, E. O., Wong, C. K., Graim, K., Bivol, A., Wang, H., Zhu, F., Afsari, B., Danilova, L. V., Favorov, A. V., Lee, W. S., Taylor, D., Hu, C. W., Long, B. L., Noren, D. P., Bisberg, A. J., , H. P. N.-D. R. E. A. M. C., Mills, G. B., Gray, J. W., Kellen, M., Norman, T., Friend, S., Qutub, A. A., Fertig, E. J., Guan, Y., Song, M., Stuart, J. M., Spellman, P. T., Koeppl, H., Stolovitzky, G., Saez-Rodriguez, J., and Mukherjee, S. (2016). Inferring causal molecular networks: empirical assessment through a community-based effort. *Nature Methods*, 13(4):310–318.
- Hodes, R. J., Sierra, F., Austad, S. N., Epel, E., Neigh, G. N., Erlandson, K. M., Schafer, M. J., LeBrasseur, N. K., Wiley, C., Campisi, J., Sehl, M. E., Scalia, R., Eguchi, S., Kasinath, B. S., Halter, J. B., Cohen, H. J., Demark-Wahnefried, W., Ahles, T. A., Barzilai, N., Hurria, A., and Hunt, P. W. (2016). Disease drivers of aging. *Annals of the New York Academy of Sciences*, 1386:45–68.
- Holzhütter, H.-G. (2004). The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry*, 271(14):2905–2922.
- Honda, Y., Tanaka, M., and Honda, S. (2010). Trehalose extends longevity in the nematode *Caenorhabditis elegans*. *Aging Cell*, 9:558–569.
- Hou, N. S. and Taubert, S. (2012). Function and regulation of lipid biology in *Caenorhabditis elegans* aging. *Frontiers in Physiology*, 3:143.
- Houthoofd, K., Braeckman, B. P., Lenaerts, I., Brys, K., Matthijssens, F., De Vreese, A., Van Eygen, S., and Vanfleteren, J. R. (2005). DAF-2 pathway mutations and food restriction in aging *Caenorhabditis elegans* differentially affect metabolism. *Neurobiology of Aging*, 26:689–696.
- Houtkooper, R. H., Mouchiroud, L., Ryu, D., Moullan, N., Katsyuba, E., Knott, G., Williams, R. W., and Auwerx, J. (2013). Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature*, 497(7450):451–457.
- Hsin, H. and Kenyon, C. (1999). Signals from the reproductive system regulate the lifespan of *C. elegans*. *Nature*, 399:362–366.
- Hsu, A.-L., Feng, Z., Hsieh, M.-Y., and Xu, X. Z. S. (2009). Identification by machine vision of the rate of motor activity decline as a lifespan predictor in *C. elegans*. *Neurobiology of Aging*, 30:1498–1503.
- Hucka, M., Finney, A., Sauro, H., Bolouri, H., Doyle, J., Kitano, H., Arkin, A., Bornstein, B., Bray, D., Cornish-Bowden, A., Cuellar, A., Dronov, S., Gilles, E., Ginkel, M., Gor, V., Goryanin, I., Hedley, W., Hodgman, T., Hofmeyr, J.-H., Hunter, P., Juty, N., Kasberger, J., Kremling, A., Kummer, U., Le Novère, N., Loew, L., Lucio, D., Mendes, P., Minch, E., Mjolsness, E., Nakayama, Y., Nelson, M., Nielsen, P., Sakurada, T., Schaff, J., Shapiro, B., Shimizu, T., Spence, H., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J., and , S. F. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531.
- Hühne, R., Thalheim, T., and Sühnel, J. (2014). AgeFactDB—the JenAge ageing factor database—towards data integration in ageing research. *Nucleic Acids Research*, 42:D892–D896.
- Ingram, D. K. and Roth, G. S. (2011). Glycolytic inhibition as a strategy for developing calorie restriction mimetics. *Experimental Gerontology*, 46(2-3):148–154.
- Ingram, D. K. and Roth, G. S. (2015). Calorie restriction mimetics: can you have your cake and eat it, too? *Ageing Research Reviews*, 20:46–62.

- Ingram, D. K., Zhu, M., Mamczarz, J., Zou, S., Lane, M. A., Roth, G. S., and deCabo, R. (2006). Calorie restriction mimetics: an emerging research field. *Aging Cell*, 5(2):97–108.
- Jensen, P. A. and Papin, J. A. (2011). Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, 27(4):541–547.
- Jensen, P. A. and Papin, J. A. (2014). MetDraw: automated visualization of genome-scale metabolic network reconstructions and high-throughput data. *Bioinformatics*, 30:1327–1328.
- Johnson, T. E. (2013). 25 years after age-1: genes, interventions and the revolution in aging research. *Experimental Gerontology*, 48:640–643.
- Kaletsky, R. and Murphy, C. T. (2010). The role of insulin/IGF-like signaling in *C. elegans* longevity and aging. *Disease Models & Mechanisms*, 3(7-8):415–419.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45:D353–D361.
- Katada, S., Imhof, A., and Sassone-Corsi, P. (2012). Connecting threads: epigenetics and metabolism. *Cell*, 148(1-2):24–28.
- Kecklund, G. and Axelsson, J. (2016). Health consequences of shift work and insufficient sleep. *BMJ Clinical Research Edition*, 355:i5210.
- Kemp, B. J., Church, D. L., Hatzold, J., Conradt, B., and Lambie, E. J. (2009). Gem-1 encodes an SLC16 monocarboxylate transporter-related protein that functions in parallel to the gon-2 TRPM channel during gonad development in *Caenorhabditis elegans*. *Genetics*, 181:581–591.
- Kenyon, C. (2011). The first long-lived mutants: discovery of the insulin/IGF-1 pathway for ageing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1561):9–16.
- Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A *C. elegans* mutant that lives twice as long as wild type. *Nature*, 366:461–464.
- King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015). Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Computational Biology*, 11(8):e1004321.
- King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis, N. E. (2016). BiGG models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44:D515–D522.
- Kirkwood, T. B. L. (2008). A systematic look at an old problem. *Nature*, 451:644–647.
- Kirkwood, T. B. L. (2011). Systems biology of ageing and longevity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366:64–70.
- Klass, M. R. (1977). Aging in the nematode *Caenorhabditis elegans*: major biological and environmental factors influencing life span. *Mechanisms of Ageing and Development*, 6:413–429.
- Klassen, A., Faccio, A. T., Canuto, G. A. B., da Cruz, P. L. R., Ribeiro, H. C., Tavares, M. F. M., and Sussulini, A. (2017). Metabolomics: Definitions and significance in systems biology. *Advances in Experimental Medicine and Biology*, 965:3–17.
- Kleessen, S., Irgang, S., Klie, S., Giavalisco, P., and Nikoloski, Z. (2015). Integration of transcriptomics and metabolomics data specifies the metabolic response of *Chlamydomonas* to rapamycin treatment. *Plant Journal*, 81(5):822–835.
- Krumsiek, J., Bartel, J., and Theis, F. J. (2016). Computational approaches for systems metabolomics. *Current Opinion in Biotechnology*, 39:198–206.
- Langmesser, S. and Albrecht, U. (2006). Life time – circadian clocks, mitochondria and metabolism. *Chronobiology International*, 23:151–157.
- Lapierre, L. R. and Hansen, M. (2012). Lessons from *C. elegans*: signaling pathways for longevity. *Trends in Endocrinology and Metabolism*, 23:637–644.
- Lee, R. Y. N., Howe, K. L., Harris, T. W., Arnaboldi, V., Cain, S., Chan, J., Chen, W. J., Davis, P., Gao, S., Grove, C., Kishore, R., Muller, H.-M., Nakamura, C., Nuin, P., Paulini, M., Raciti, D., Rodgers, F., Russell, M., Schindelman, G., Tuli, M. A., Van Auken, K., Wang, Q., Williams, G., Wright, A., Yook, K., Berriman, M., Kersey, P., Schedl, T., Stein, L., and Sternberg, P. W. (2018). WormBase 2017: molting into a new stage. *Nucleic Acids Research*, 46:D869–D874.
- Lemieux, G. A. and Ashrafi, K. (2016). Investigating connections between metabolism, longevity, and behavior in *Caenorhabditis elegans*. *Trends in Endocrinology and Metabolism*, 27:586–596.
- Lewis, N. E., Nagarajan, H., and Palsson, B. O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305.
- Liu, H., Guo, M., Xue, T., Guan, J., Luo, L., and Zhuang, Z. (2016). Screening lifespan-extending drugs in *Caenorhabditis elegans* via label propagation on drug-protein networks. *BMC Systems Biology*, 10:131.

- Liu, X. and Locasale, J. W. (2017). Metabolomics: A primer. *Trends in Biochemical Sciences*, 42:274–284.
- Lu, N. and Balachandar, R. (2005). Nutritional requirements for pantothenate, pantethine or coenzyme A in the free-living nematode *Caenorhabditis elegans*. *Nematology*, 7(5):761–766.
- Lubitz, T., Hahn, J., Bergmann, F. T., Noor, E., Klipp, E., and Liebermeister, W. (2016). SBtab: a flexible table format for data exchange in systems biology. *Bioinformatics*, 32:2559–2561.
- Lund, J., Tedesco, P., Duke, K., Wang, J., Kim, S. K., and Johnson, T. E. (2002). Transcriptional profile of aging in *C. elegans*. *Current Biology*, 12:1566–1573.
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell*, 153:1194–1217.
- Ma, L., Chan, A. H. C., Hattwell, J. P. N., Ebert, P. R., and Schirra, H. J. (2017). Systems biology analysis using a genome-scale metabolic model shows that phosphine triggers global metabolic suppression in a resistant strain of *C. elegans*. *BioRxiv*, doi: <https://doi.org/10.1101/144386>.
- Maarleveld, T. R., Khandelwal, R. A., Olivier, B. G., Teusink, B., and Bruggeman, F. J. (2013). Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology*, 8:997–1008.
- Machado, D. and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Computational Biology*, 10(4):e1003580.
- MacNeil, L. T., Watson, E., Arda, H. E., Zhu, L. J., and Walhout, A. J. M. (2013). Diet-induced developmental acceleration independent of TOR and insulin in *C. elegans*. *Cell*, 153(1):240–252.
- Maglioni, S., Schiavi, A., Runci, A., Shaik, A., and Ventura, N. (2014). Mitochondrial stress extends lifespan in *C. elegans* through neuronal hormesis. *Experimental Gerontology*, 56:89–98.
- Mains, A. (2018). *Germline loss dynamically remodels the metabolic transcriptional network over young-adulthood. A time-resolved analysis of transcriptional changes throughout adulthood of C. elegans with differing gonadal composition*. PhD thesis, University of Cambridge.
- Mair, W., Morantte, I., Rodrigues, A. P. C., Manning, G., Montminy, M., Shaw, R. J., and Dillin, A. (2011). Lifespan extension induced by AMPK and calcineurin is mediated by CRTC-1 and CREB. *Nature*, 470:404–408.
- Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., , D. R. E. A. M. C., Kellis, M., Collins, J. J., and Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8):796–804.
- Mardinoglu, A. and Nielsen, J. (2012). Systems medicine and metabolic modelling. *Journal of Internal Medicine*, 271:142–154.
- Mathew, R., Pal Bhadra, M., and Bhadra, U. (2017). Insulin/insulin-like growth factor-1 signalling (IIS) based regulation of lifespan across species. *Biogerontology*, 18:35–53.
- McCormick, M., Chen, K., Ramaswamy, P., and Kenyon, C. (2012). New genes that extend *Caenorhabditis elegans* lifespan in response to reproductive signals. *Aging Cell*, 11:192–202.
- McCormick, M. A. and Promislow, D. E. L. (2018). Recent advances in the systems biology of aging. *Antioxidants & Redox Signaling*, 29:973–984.
- McMurry, J. A., Juty, N., Blomberg, N., Burdett, T., Conlin, T., Conte, N., Courtot, M., Deck, J., Dumontier, M., Fellows, D. K., Gonzalez-Beltran, A., Gormanns, P., Grethe, J., Hastings, J., Hériché, J.-K., Hermjakob, H., Ison, J. C., Jimenez, R. C., Jupp, S., Kunze, J., Laibe, C., Le Novère, N., Malone, J., Martin, M. J., McEntyre, J. R., Morris, C., Muilu, J., Müller, W., Rocca-Serra, P., Sansone, S.-A., Sariyar, M., Snoep, J. L., Soiland-Reyes, S., Stanford, N. J., Swainston, N., Washington, N., Williams, A. R., Wimalaratne, S. M., Winfree, L. M., Wolstencroft, K., Goble, C., Mungall, C. J., Haendel, M. A., and Parkinson, H. (2017). Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biology*, 15:e2001414.
- Megchelenbrink, W., Huynen, M., and Marchiori, E. (2014). optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS One*, 9:e86587.
- Migliori, M. L., Simonetta, S. H., Romanowski, A., and Golombek, D. A. (2011). Circadian rhythms in metabolic variables in *Caenorhabditis elegans*. *Physiology & Behavior*, 103:315–320.
- Minois, N. (2014). Molecular basis of the ‘anti-aging’ effect of spermidine and other natural polyamines - a mini-review. *Gerontology*, 60:319–326.
- Minois, N., Rockenfeller, P., Smith, T. K., and Carmona-Gutierrez, D. (2014). Spermidine feeding decreases age-related locomotor activity loss and induces changes in lipid composition. *PLoS One*, 9:e102435.
- Mishur, R. J., Khan, M., Munkácsy, E., Sharma, L., Bokov, A., Beam, H., Radetskaya, O., Borror, M., Lane, R., Bai, Y., and Rea, S. L. (2016). Mitochondrial metabolites extend lifespan. *Aging Cell*, 15:336–348.
- Monteverde, T., Muthalagu, N., Port, J., and Murphy, D. J. (2015). Evidence of cancer-promoting roles for AMPK and related kinases. *The FEBS Journal*, 282:4658–4671.

- Morcos, M., Du, X., Pfisterer, F., Hutter, H., Sayed, A. A. R., Thornalley, P., Ahmed, N., Baynes, J., Thorpe, S., Kukudov, G., Schlotterer, A., Bozorgmehr, F., El Baki, R. A., Stern, D., Moehrlen, F., Ibrahim, Y., Oikonomou, D., Hamann, A., Becker, C., Zeier, M., Schwenger, V., Miftari, N., Humpert, P., Hammes, H.-P., Buechler, M., Bierhaus, A., Brownlee, M., and Nawroth, P. P. (2008). Glyoxalase-1 prevents mitochondrial protein modification and enhances lifespan in *Caenorhabditis elegans*. *Aging Cell*, 7:260–269.
- Moreno-Arriola, E., El Hafidi, M., Ortega-Cuellar, D., and Carvajal, K. (2016). AMP-activated protein kinase regulates oxidative metabolism in *Caenorhabditis elegans* through the NHR-49 and MDT-15 transcriptional regulators. *PLoS One*, 11:e0148089.
- Moretti, S., Martin, O., Van Du Tran, T., Bridge, A., Morgat, A., and Pagni, M. (2016). MetaNetX/MNXref—reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Research*, 44:D523–D526.
- Morgan, A. E., Mooney, K. M., Wilkinson, S. J., Pickles, N. A., and Mc Auley, M. T. (2016). Cholesterol metabolism: A review of how ageing disrupts the biological mechanisms responsible for its regulation. *Ageing Research Reviews*, 27:108–124.
- Mouchiroud, L., Houtkooper, R. H., Moullan, N., Katsyuba, E., Ryu, D., Cantó, C., Mottis, A., Jo, Y.-S., Viswanathan, M., Schoonjans, K., Guarente, L., and Auwerx, J. (2013). The NAD(+)/sirtuin pathway modulates longevity through activation of mitochondrial UPR and FOXO signaling. *Cell*, 154:430–441.
- Nakahata, Y., Sahar, S., Astarita, G., Kaluzova, M., and Sassone-Corsi, P. (2009). Circadian control of the NAD+ salvage pathway by CLOCK-SIRT1. *Science*, 324:654–657.
- Niccoli, T. and Partridge, L. (2012). Ageing as a risk factor for disease. *Current Biology*, 22(17):R741–R752.
- Omodei, D. and Fontana, L. (2011). Calorie restriction and prevention of age-associated chronic disease. *FEBS Letters*, 585(11):1537–1542.
- Onken, B. and Driscoll, M. (2010). Metformin induces a dietary restriction-like state and the oxidative stress response to extend *C. elegans* healthspan via AMPK, LKB1, and SKN-1. *PLoS One*, 5(1):e8758.
- O'Rourke, E. J., Kuballa, P., Xavier, R., and Ruvkun, G. (2013). ω -6 polyunsaturated fatty acids extend life span through the activation of autophagy. *Genes & Development*, 27:429–440.
- O'Rourke, E. J., Soukas, A. A., Carr, C. E., and Ruvkun, G. (2009). *C. elegans* major fats are stored in vesicles distinct from lysosome-related organelles. *Cell Metabolism*, 10:430–435.
- Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., and Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology*, 7:535.
- Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248.
- Panowski, S. H. and Dillin, A. (2009). Signals of youth: endocrine regulation of aging in *Caenorhabditis elegans*. *Trends in Endocrinology and Metabolism*, 20:259–264.
- Partridge, L., Deelen, J., and Slagboom, P. E. (2018). Facing up to the global challenges of ageing. *Nature*, 561:45–56.
- Partridge, L., Gems, D., and Withers, D. J. (2005). Sex and death: what is the connection? *Cell*, 120:461–472.
- Pfau, T., Christian, N., and Ebenhöf, O. (2011). Systems approaches to modelling pathways and networks. *Briefings in Functional Genomics*, 10:266–279.
- Pinto, R. C., Trygg, J., and Gottfries, J. (2012). Advantages of orthogonal inspection in chemometrics. *Journal of Chemometrics*, 26:231–235.
- Pontoizeau, C., Mouchiroud, L., Molin, L., Mergoud-Dit-Lamarche, A., Dallièrre, N., Toulhoat, P., Elena-Herrmann, B., and Solari, F. (2014). Metabolomics analysis uncovers that dietary restriction buffers metabolic changes associated with aging in *Caenorhabditis elegans*. *Journal of Proteome Research*, 13:2910–2919.
- Pu, M., Ni, Z., Wang, M., Wang, X., Wood, J. G., Helfand, S. L., Yu, H., and Lee, S. S. (2015). Trimethylation of Lys36 on H3 restricts gene expression change during aging and impacts life span. *Genes & Development*, 29(7):718–731.
- Pu, M., Wang, M., Wang, W., Velayudhan, S. S., and Lee, S. S. (2018). Unique patterns of trimethylation of histone H3 lysine 4 are prone to changes during aging in *Caenorhabditis elegans* somatic cells. *PLoS Genetics*, 14:e1007466.
- Putri, S. P., Nakayama, Y., Matsuda, F., Uchikata, T., Kobayashi, S., Matsubara, A., and Fukusaki, E. (2013). Current metabolomics: practical applications. *Journal of Bioscience and Bioengineering*, 115:579–589.
- Qi, W., Gutierrez, G. E., Gao, X., Dixon, H., McDonough, J. A., Marini, A. M., and Fisher, A. L. (2017). The ω -3 fatty acid α -linolenic acid extends *Caenorhabditis elegans* lifespan via NHR-49/PPARA and oxidation to oxylipins. *Aging Cell*, 16:1125–1135.
- Quach, A., Levine, M. E., Tanaka, T., Lu, A. T., Chen, B. H., Ferrucci, L., Ritz, B., Bandinelli, S., Neuhauser, M. L., Beasley, J. M., Snetselaar, L., Wallace, R. B., Tsao, P. S., Absher, D., Assimes, T. L., Stewart, J. D., Li, Y.,

- Hou, L., Baccarelli, A. A., Whitsel, E. A., and Horvath, S. (2017). Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging*, 9:419–446.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, A. U., Carta, L. K., Lesuisse, E., and Hamza, I. (2005). Lack of heme synthesis in a free-living eukaryote. *Proceedings of the National Academy of Sciences of the United States of America*, 102:4270–4275.
- Rea, I. M. (2017). Towards ageing well: Use it or lose it: Exercise, epigenetics and cognition. *Biogerontology*, 18:679–691.
- Regmi, S. G., Rolland, S. G., and Conradt, B. (2014). Age-dependent changes in mitochondrial morphology and volume are not predictors of lifespan. *Aging*, 6:118–130.
- Revelle, W. (2018). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 1.8.3.
- Reznik, E., Mehta, P., and Segrè, D. (2013). Flux imbalance analysis and the sensitivity of cellular growth to changes in metabolite pools. *PLoS Computational Biology*, 9:e1003195.
- Riera, C. E. and Dillin, A. (2016). Emerging role of sensory perception in aging and metabolism. *Trends in Endocrinology and Metabolism*, 27(5):294–303.
- Ristow, M. and Zarse, K. (2010). How increased oxidative stress promotes longevity and metabolic health: The concept of mitochondrial hormesis (mitohormesis). *Experimental Gerontology*, 45:410–418.
- Rosato, A., Tenori, L., Cascante, M., De Atauri Carulla, P. R., Martins Dos Santos, V. A. P., and Saccenti, E. (2018). From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics*, 14:37.
- Rossum, G. (1995). Python reference manual. Technical report, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.
- Roth, G. S. and Ingram, D. K. (2016). Manipulation of health span and function by dietary caloric restriction mimetics. *Annals of the New York Academy of Sciences*, 1363:5–10.
- Russell, J. C., Burnaevskiy, N., Ma, B., Mailig, M. A., Faust, F., Crane, M., Kaeberlein, M., and Mendenhall, A. (2017). Electrophysiological measures of aging pharynx function in *C. elegans* reveal enhanced organ functionality in older, long-lived mutants. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, page glx230.
- Savage, M. O. (2013). Phenotypes, investigation and treatment of primary IGF-1 deficiency. *Endocrine Development*, 24:138–149.
- Schmidt, B. J., Ebrahim, A., Metz, T. O., Adkins, J. N., Palsson, B. Ø., and Hyduke, D. R. (2013). GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics*, 29(22):2900–2908.
- Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3:119.
- Schulz, T. J., Zarse, K., Voigt, A., Urban, N., Birringer, M., and Ristow, M. (2007). Glucose restriction extends *Caenorhabditis elegans* life span by inducing mitochondrial respiration and increasing oxidative stress. *Cell Metabolism*, 6:280–293.
- Seo, Y., Kingsley, S., Walker, G., Mondoux, M. A., and Tissenbaum, H. A. (2018). Metabolic shift from glycogen to trehalose promotes lifespan and healthspan *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, 115:E2791–E2800.
- Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø., and Rupp, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nature Biotechnology*, 26(9):1003–1010.
- Simonetta, S. H., Migliori, M. L., Romanowski, A., and Golombek, D. A. (2009). Timing of locomotor activity circadian rhythms in *Caenorhabditis elegans*. *PLoS One*, 4:e7571.
- Simonetta, S. H., Romanowski, A., Minniti, A. N., Inestrosa, N. C., and Golombek, D. A. (2008). Circadian stress tolerance in adult *Caenorhabditis elegans*. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 194:821–828.
- Smith, A. C., Eyassu, F., Mazat, J.-P., and Robinson, A. J. (2017). MitoCore: a curated constraint-based model for simulating human central metabolism. *BMC Systems Biology*, 11:114.
- Soltow, Q. A., Jones, D. P., and Promislow, D. E. L. (2010). A network perspective on metabolism and aging. *Integrative and Comparative Biology*, 50:844–854.
- Soultoukis, G. A. and Partridge, L. (2016). Dietary protein, metabolism, and aging. *Annual Review of Biochemistry*, 85:5–34.
- Steinbaugh, M. J., Narasimhan, S. D., Robida-Stubbs, S., Moronetti Mazzeo, L. E., Dreyfuss, J. M., Hourihan, J. M., Raghavan, P., Operaña, T. N., Esmailie, R., and Blackwell, T. K. (2015). Lipid-mediated regulation of

- SKN-1/Nrf in response to germ cell absence. *Elife*, 4.
- Stroustrup, N., Ulmschneider, B. E., Nash, Z. M., López-Moyado, I. F., Apfeld, J., and Fontana, W. (2013). The *Caenorhabditis elegans* lifespan machine. *Nature Methods*, 10:665–670.
- Su, G., Morris, J. H., Demchak, B., and Bader, G. D. (2014). Biological network exploration with Cytoscape 3. *Current Protocols in Bioinformatics*, 47:8.13.1–8.13.24.
- Sun, A. Y. and Lambie, E. J. (1997). gon-2, a gene required for gonadogenesis in *Caenorhabditis elegans*. *Genetics*, 147:1077–1089.
- Swainston, N., Smallbone, K., Hefzi, H., Dobson, P. D., Brewer, J., Hanscho, M., Zielinski, D. C., Ang, K. S., Gardiner, N. J., Gutierrez, J. M., Kyriakopoulos, S., Lakshmanan, M., Li, S., Liu, J. K., Martínez, V. S., Orellana, C. A., Quek, L.-E., Thomas, A., Zanghellini, J., Borth, N., Lee, D.-Y., Nielsen, L. K., Kell, D. B., Lewis, N. E., and Mendes, P. (2016). Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*, 12(7):1–7.
- Sweetlove, L. J., Obata, T., and Fernie, A. R. (2014). Systems analysis of metabolic phenotypes: what have we learnt? *Trends in Plant Science*, 19:222–230.
- Swire, J., Fuchs, S., Bundy, J. G., and Leroi, A. M. (2009). The cellular geometry of growth drives the amino acid economy of *Caenorhabditis elegans*. *Proceedings: Biological Sciences*, 276:2747–2754.
- The Gene Ontology Consortium (2019). The Gene Ontology resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47:D330–D338.
- Thiele, I. and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1):93–121.
- Thiele, I., Swainston, N., Fleming, R. M. T., Hoppe, A., Sahoo, S., Aurich, M. K., Haraldsdottir, H., Mo, M. L., Rolfsson, O., Stobbe, M. D., Thorleifsson, S. G., Agren, R., Bölling, C., Bordel, S., Chavali, A. K., Dobson, P., Dunn, W. B., Endler, L., Hala, D., Hucka, M., Hull, D., Jameson, D., Jamshidi, N., Jonsson, J. J., Juty, N., Keating, S., Nookaew, I., Le Novère, N., Malys, N., Mazein, A., Papin, J. A., Price, N. D., Selkov, Sr, E., Sigurdsson, M. I., Simeonidis, E., Sonnenschein, N., Smallbone, K., Sorokin, A., van Beek, J. H. G. M., Weichart, D., Goryanin, I., Nielsen, J., Westerhoff, H. V., Kell, D. B., Mendes, P., and Palsson, B. Ø. (2013). A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, 31(5):419–425.
- Thondamal, M., Witting, M., Schmitt-Kopplin, P., and Aguilaniu, H. (2014). Steroid hormone signalling links reproduction to lifespan in dietary-restricted *Caenorhabditis elegans*. *Nature Communications*, 5:4879.
- Thévenot, E. A., Roux, A., Xu, Y., Ezan, E., and Junot, C. (2015). Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *Journal of Proteome Research*, 14:3322–3335.
- Tian, Y., Garcia, G., Bian, Q., Steffen, K. K., Joe, L., Wolff, S., Meyer, B. J., and Dillin, A. (2016). Mitochondrial stress induces chromatin reorganization to promote longevity and UPR(mt). *Cell*, 165:1197–1208.
- Timmerman, K. L. and Volpi, E. (2010). Amino acid metabolism and regulatory effects in aging. *Current Opinion in Clinical Nutrition & Metabolic Care*, 11(1):45–49.
- Tissenbaum, H. A. and Guarente, L. (2001). Increased dosage of a sir-2 gene extends lifespan in *Caenorhabditis elegans*. *Nature*, 410:227–230.
- Tissenbaum, H. A. and Guarente, L. (2002). Model organisms as a guide to mammalian aging. *Developmental Cell*, 2:9–19.
- Töpfer, N., Kleessen, S., and Nikoloski, Z. (2015). Integration of metabolomics data into metabolic networks. *Frontiers in Plant Science*, 6:49.
- Torgovnick, A., Schiavi, A., Maglioni, S., and Ventura, N. (2013). Healthy aging: what can we learn from *Caenorhabditis elegans*? *Zeitschrift für Gerontologie und Geriatrie*, 46:623–628.
- Tretter, L., Patocs, A., and Chinopoulos, C. (2016). Succinate, an intermediate in metabolism, signal transduction, ROS, hypoxia, and tumorigenesis. *Biochimica et Biophysica Acta*, 1857:1086–1101.
- Tumminia, A., Vinciguerra, F., Parisi, M., and Frittitta, L. (2018). Type 2 diabetes mellitus and alzheimer's disease: Role of insulin signalling and therapeutic implications. *International Journal of Molecular Sciences*, 19.
- Töpfer, N., Scossa, F., Fernie, A., and Nikoloski, Z. (2014). Variability of metabolite levels is linked to differential metabolic pathways in *Arabidopsis*'s responses to abiotic stresses. *PLoS Computational Biology*, 10:e1003656.
- Töpfer, N., Seaver, S. M. D., and Aharoni, A. (2018). Integration of plant metabolomics data with metabolic networks: Progresses and challenges. *Methods in Molecular Biology*, 1778:297–310.
- Urban, N., Tsitsipatis, D., Hausig, F., Kreuzer, K., Erler, K., Stein, V., Ristow, M., Steinbrenner, H., and Klotz, L.-O. (2017). Non-linear impact of glutathione depletion on *C. elegans* life span and stress resistance. *Redox Biology*, 11:502–515.
- Valdes, A. M., Glass, D., and Spector, T. D. (2013). Omics technologies and the study of human ageing. *Nature Reviews Genetics*, 14:601–607.

- Van Assche, R., Broeckx, V., Boonen, K., Maes, E., De Haes, W., Schoofs, L., and Temmerman, L. (2015). Integrating -omics: Systems biology as explored through *C. elegans* research. *Journal of Molecular Biology*, 427:3441–3451.
- van Delft, P., Akay, A., Huber, S. M., Bueschl, C., Rudolph, K. L. M., Di Domenico, T., Schuhmacher, R., Miska, E. A., and Balasubramanian, S. (2017). The profile and dynamics of RNA modifications in animals. *ChemBioChem*, 18:979–984.
- van der Linden, A. M., Beverly, M., Kadener, S., Rodriguez, J., Wasserman, S., Rosbash, M., and Sengupta, P. (2010). Genome-wide analysis of light- and temperature-entrained circadian transcripts in *Caenorhabditis elegans*. *PLoS Biology*, 8:e1000503.
- Veech, R. L., Bradshaw, P. C., Clarke, K., Curtis, W., Pawlosky, R., and King, M. T. (2017). Ketone bodies mimic the life span extending properties of caloric restriction. *IUBMB Life*, 69:305–314.
- Vellai, T., Takacs-Vellai, K., Zhang, Y., Kovacs, A. L., Orosz, L., and Müller, F. (2003). Genetics: influence of TOR kinase on lifespan in *C. elegans*. *Nature*, 426:620.
- Vergano, S. S., Rao, M., McCormack, S., Ostrovsky, J., Clarke, C., Preston, J., Bennett, M. J., Yudkoff, M., Xiao, R., and Falk, M. J. (2014). In vivo metabolic flux profiling with stable isotopes discriminates sites and quantifies effects of mitochondrial dysfunction in *C. elegans*. *Molecular Genetics and Metabolism*, 111:331–341.
- Villaverde, A. F. and Banga, J. R. (2014). Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of the Royal Society Interface*, 11(91):20130505.
- Virk, B., Correia, G., Dixon, D. P., Feyst, I., Jia, J., Oberleitner, N., Briggs, Z., Hodge, E., Edwards, R., Ward, J., Gems, D., and Weinkove, D. (2012). Excessive folate synthesis limits lifespan in the *C. elegans*: *E. coli* aging model. *BMC Biology*, 10:67.
- Virk, B., Jia, J., Maynard, C. A., Raimundo, A., Lefebvre, J., Richards, S. A., Chetina, N., Liang, Y., Helliwell, N., Cipinska, M., and Weinkove, D. (2016). Folate acts in *E. coli* to accelerate *C. elegans* aging independently of bacterial biosynthesis. *Cell Reports*, 14(7):1611–1620.
- Wan, Q.-L., Shi, X., Liu, J., Ding, A.-J., Pu, Y.-Z., Li, Z., Wu, G.-S., and Luo, H.-R. (2017). Metabolomic signature associated with reproduction-regulated aging in *Caenorhabditis elegans*. *Aging*, 9:447–474.
- Wang, T. J., Ngo, D., Psychogios, N., Dejam, A., Larson, M. G., Vasan, R. S., Ghorbani, A., O'Sullivan, J., Cheng, S., Rhee, E. P., Sinha, S., McCabe, E., Fox, C. S., O'Donnell, C. J., Ho, J. E., Florez, J. C., Magnusson, M., Pierce, K. A., Souza, A. L., Yu, Y., Carter, C., Light, P. E., Melander, O., Clish, C. B., and Gerszten, R. E. (2013). 2-aminoadipic acid is a biomarker for diabetes risk. *The Journal of Clinical Investigation*, 123:4309–4317.
- Webster, C. M., Pino, E. C., Carr, C. E., Wu, L., Zhou, B., Cedillo, L., Kacergis, M. C., Curran, S. P., and Soukas, A. A. (2017). Genome-wide RNAi screen for fat regulatory genes in *C. elegans* identifies a proteostasis-AMPK axis critical for starvation survival. *Cell Reports*, 20:627–640.
- Weir, H. J., Yao, P., Huynh, F. K., Escoubas, C. C., Goncalves, R. L., Burkewitz, K., Laboy, R., Hirschey, M. D., and Mair, W. B. (2017). Dietary restriction and AMPK increase lifespan via mitochondrial network and peroxisome remodeling. *Cell Metabolism*, 26:884–896.e5.
- Wellen, K. E., Hatzivassiliou, G., Sachdeva, U. M., Bui, T. V., Cross, J. R., and Thompson, C. B. (2009). ATP-citrate lyase links cellular metabolism to histone acetylation. *Science*, 324:1076–1080.
- Willemsen, A. M., Hendrickx, D. M., Hoefsloot, H. C. J., Hendriks, M. M. W. B., Wahl, S. A., Teusink, B., Smilde, A. K., and van Kampen, A. H. C. (2015). MetDFBA: incorporating time-resolved metabolomics measurements into dynamic flux balance analysis. *Molecular Biosystems*, 11(1):137–145.
- Willenberg, I., Ostermann, A. I., and Schebb, N. H. (2015). Targeted metabolomics of the arachidonic acid cascade: current state and challenges of LC-MS analysis of oxylipins. *Analytical and Bioanalytical Chemistry*, 407:2675–2683.
- Williams, D. S., Cash, A., Hamadani, L., and Diemer, T. (2009). Oxaloacetate supplementation increases lifespan in *Caenorhabditis elegans* through an AMPK/FOXO-dependent pathway. *Aging Cell*, 8:765–768.
- Witting, M., Hastings, J., Rodriguez, N., Joshi, C. J., Hattwell, J. P., Ebert, P. R., van Weeghel, M., Wakelam, M., Houtkooper, R., Mains, A., Sadykoff, S., Schroeder, F., Lewis, N. E., Kaleta, C., Casanueva, O., Novère, N. L., and Schirra, H.-J. (2018). Modeling meets metabolomics – the WormJam consensus model as basis for metabolic studies in the model organism *Caenorhabditis elegans*. *Frontiers in Molecular Biosciences*, 5:96.
- Witting, M. and Schmitt-Kopplin, P. (2016). The *Caenorhabditis elegans* lipidome: A primer for lipid analysis in *Caenorhabditis elegans*. *Archives of Biochemistry and Biophysics*, 589:27–37.
- Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58:109–130.
- Xia, J. (2017). Computational strategies for biological interpretation of metabolomics data. *Advances in Experimental Medicine and Biology*, 965:191–206.

- Yamawaki, T. M., Arantes-Oliveira, N., Berman, J. R., Zhang, P., and Kenyon, C. (2008). Distinct activities of the germline and somatic reproductive tissues in the regulation of *Caenorhabditis elegans*' longevity. *Genetics*, 178:513–526.
- Yamawaki, T. M., Berman, J. R., Suchanek-Kavipurapu, M., McCormick, M., Gaglia, M. M., Lee, S.-J., and Kenyon, C. (2010). The somatic reproductive tissues of *C. elegans* promote longevity through steroid hormone signaling. *PLoS Biology*, 8(8):e1000468.
- Yang, C., Hua, Q., and Shimizu, K. (2002). Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*. *Applied Microbiology and Biotechnology*, 58(6):813–822.
- Yasuda, K., Hartman, P. S., Ishii, T., Suda, H., Akatsuka, A., Shoyama, T., Miyazawa, M., and Ishii, N. (2011). Interrelationships between mitochondrial fusion, energy metabolism and oxidative stress during development in *Caenorhabditis elegans*. *Biochemical and Biophysical Research Communications*, 404:751–755.
- Yasuda, K., Ishii, T., Suda, H., Akatsuka, A., Hartman, P. S., Goto, S., Miyazawa, M., and Ishii, N. (2006). Age-related changes of mitochondrial structure and function in *Caenorhabditis elegans*. *Mechanisms of Ageing and Development*, 127:763–770.
- Yi, H.-S., Chang, J. Y., and Shong, M. (2018). The mitochondrial unfolded protein response and mitohormesis: a perspective on metabolic diseases. *Journal of Molecular Endocrinology*, 61:R91–R105.
- Yilmaz, L. S. and Walhout, A. J. M. (2016). A *Caenorhabditis elegans* genome-scale metabolic network model. *Cell Systems*, 2(5):297–311.
- Yuan, Y., Hakimi, P., Kao, C., Kao, A., Liu, R., Janocha, A., Boyd-Tressler, A., Hang, X., Alhoraibi, H., Slater, E., Xia, K., Cao, P., Shue, Q., Ching, T.-T., Hsu, A.-L., Erzurum, S. C., Dubyak, G. R., Berger, N. A., Hanson, R. W., and Feng, Z. (2016). Reciprocal changes in phosphoenolpyruvate carboxykinase and pyruvate kinase with age are a determinant of aging in *Caenorhabditis elegans*. *The Journal of Biological Chemistry*, 291:1307–1319.
- Yugi, K., Kubota, H., Hatano, A., and Kuroda, S. (2016). Trans-Omics: How to reconstruct biochemical networks across multiple 'omic' layers. *Trends in Biotechnology*, 34:276–290.
- Yugi, K. and Kuroda, S. (2017). Metabolism-centric trans-omics. *Cell Systems*, 4:19–20.
- Yun, J. and Finkel, T. (2014). Mitohormesis. *Cell Metabolism*, 19:757–766.
- Zamboni, N., Saghatelian, A., and Patti, G. J. (2015). Defining the metabolome: size, flux, and regulation. *Molecular Cell*, 58:699–706.
- Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Research*, 33:W741–W748.
- Zhang, B., Xiao, R., Ronan, E. A., He, Y., Hsu, A.-L., Liu, J., and Xu, X. Z. S. (2015). Environmental temperature differentially modulates *C. elegans* longevity through a thermosensitive TRP channel. *Cell Reports*, 11:1414–1424.
- Zhou, K. I., Pincus, Z., and Slack, F. J. (2011). Longevity and stress in *Caenorhabditis elegans*. *Ageing*, 3:733–753.
- Zhou, Y., Wang, Y., Zhang, X., Bhar, S., Jones Lipinski, R. A., Han, J., Feng, L., and Butcher, R. A. (2018). Biosynthetic tailoring of existing ascaroside pheromones alters their biological function in *C. elegans*. *Elife*, 7.
- Zhu, B., Zhang, Q., Pan, Y., Mace, E. M., York, B., Antoulas, A. C., Dacso, C. C., and O'Malley, B. W. (2017). A cell-autonomous mammalian 12 hr clock coordinates metabolic and stress rhythms. *Cell Metabolism*, 25:1305–1319.e9.
- Zierer, J., Menni, C., Kastenmüller, G., and Spector, T. D. (2015). Integration of 'omics' data in aging research: from biomarkers to systems biology. *Ageing Cell*, 14:933–944.
- Zur, H., Ruppin, E., and Shlomi, T. (2010). iMAT: an integrative metabolic analysis tool. *Bioinformatics*, 26(24):3140–3142.

Appendix A. Supplementary Tables And Figures

Supplementary Tables

Table A.1 Names, pathways and additional information for measured metabolites in metabolomics assay

| Name | Mass / RT | Pathway | Formula | ChEBI ID |
|---------------------|------------------|-----------------------------------|-----------|-------------|
| Pyruvate | 87.0 / 43.0 | Glycolysis | C3H3O3 | CHEBI:15361 |
| Lactate | 89.0 / 43.0 | Glycolysis; TCA cycle | C3H5O3 | CHEBI:24996 |
| Oxaloacetate | 131.0 / 113.0 | Glycolysis; TCA cycle | C4H2O5 | CHEBI:16452 |
| PEP | 167.0 / 79.0 | Glycolysis | C3H5O6P | CHEBI:44897 |
| DHAP | 167.0 / 79.0 (2) | Glycolysis; Gluconeogenesis | C3H7O6P | CHEBI:16108 |
| D-GA3P | 167.0 / 97.0 | Glycolysis; Gluconeogenesis | C3H5O6P | CHEBI:59776 |
| Glucose | 179.0 / 89.0 | Glycolysis; Gluconeogenesis | C6H12O6 | CHEBI:17234 |
| G1P/G6P/F6P/F1P | 259.0 / 97.0 | Glycolysis | C6H13O9P | CHEBI:29042 |
| 3HBA | 103.0 / 59.0 | TCA cycle | C4H8O3 | CHEBI:20067 |
| Maleic Acid | 115.0 / 71.0 (2) | TCA cycle | C4H4O4 | CHEBI:18300 |
| Succinate | 117.0 / 73.0 | TCA cycle | C4H5O4 | CHEBI:26806 |
| Malate | 133.0 / 115.0 | TCA cycle | C4H4O5 | CHEBI:25115 |
| Aconitate | 173.0 / 85.0 | TCA cycle | C6H3O6 | CHEBI:22210 |
| Glycine | 76.0 / 30.1 | Amino acid | C2H5NO2 | CHEBI:15428 |
| Alanine | 90.0 / 44.0 (2) | Amino acid | C3H7NO2 | CHEBI:16449 |
| Cadaverine | 103.0 / 86.0 | Amine | C5H14N2 | CHEBI:18127 |
| Serine | 106.0 / 60.0 (2) | Amino acid | C3H7NO3 | CHEBI:17822 |
| Proline | 116.0 / 70.0 | Amino acid | C5H9NO2 | CHEBI:26271 |
| Valine | 118.0 / 72.0 | Amino acid | C5H11NO2 | CHEBI:27266 |
| Threonine | 120.0 / 74.0 (2) | Amino acid | C4H9NO3 | CHEBI:26986 |
| Leucine | 132.0 / 86.0 | Amino acid | C6H13NO2 | CHEBI:25017 |
| iso-Leucine | 132.0 / 86.0 (2) | Amino acid | C6H13NO2 | CHEBI:24898 |
| Asparagine | 133.0 / 74.0 | Amino acid | C4H8N2O3 | CHEBI:22653 |
| Tyramine | 138.0 / 121.0 | Amine | C8H11NO | CHEBI:15760 |
| Glutamine | 147.0 / 84.0 | Amino acid | C5H10N2O3 | CHEBI:28300 |
| Lysine | 147.0 / 84.0 (2) | Amino acid | C6H14N2O2 | CHEBI:25094 |
| Glutamic Acid | 148.0 / 84.0 | Amino acid | C5H9NO4 | CHEBI:18237 |
| Methionine | 150.0 / 61.0 | Amino acid | C5H11NO2S | CHEBI:16811 |
| Histidine | 156.0 / 110.0 | Amino acid | C6H9N3O2 | CHEBI:27570 |
| Tryptamine | 161.0 / 144.0 | Amino acid | C10H12N2 | CHEBI:16765 |
| Phenylalanine | 166.0 / 120.0 | Amino acid | C9H11NO2 | CHEBI:28044 |
| 1/3-Methylhistidine | 170.0 / 96.0 | Amino acid (Histidine Metabolism) | C7H11N3O2 | CHEBI:70958 |
| Arginine | 175.0 / 70.0 | Amino acid | C6H14N4O2 | CHEBI:29016 |

Continuation of Table A.1

| Name | Mass / RT | Pathway | Formula | ChEBI ID |
|----------------------------|---------------|---|---|--------------|
| Tyrosine | 182.1 / 136.0 | Amino Acid | C ₉ H ₁₁ NO ₃ | CHEBI:18186 |
| Cystine | 241.1 / 120.0 | Amino Acid | C ₆ H ₁₂ N ₂ O ₄ S ₂ | CHEBI:17376 |
| 2-Hydroxyisovaleric Acid | 117.0 / 71.0 | Amino Acid | C ₅ H ₁₀ O ₃ | CHEBI:60645 |
| 3-Methyl-2-Oxovaleric Acid | 129.0 / 101.0 | Amino Acid | C ₆ H ₁₀ O ₃ | CHEBI:35932 |
| Aminoisobutyrate | 104.0 / 86.0 | Amino acids metabolism/Val, Leu, iso-Leu | C ₄ H ₈ NO ₂ | CHEBI:49096 |
| Betaine | 118.0 / 58.0 | Amino acids metabolism/Gly,Ser, Thr metabolism | C ₅ H ₁₁ NO ₂ | CHEBI:17750 |
| Homoserine | 120.0 / 74.0 | Amino acids metabolism/Thr, Met, Asp | C ₄ H ₉ NO ₃ | CHEBI:30653 |
| Creatine | 132.0 / 90.0 | Amino acids metabolism/Arg, Gly | C ₄ H ₉ N ₃ O ₂ | CHEBI:16919 |
| Aspartic Acid | 134.0 / 74.0 | Amino Acid | C ₄ H ₇ NO ₄ | CHEBI:22660 |
| Carnitine | 162.0 / 85.0 | Amino acids metabolism/Lys | C ₇ H ₁₅ NO ₃ | CHEBI:17126 |
| Quinolinic Acid | 168.0 / 150.0 | Amino Acid/ Tryptophan, beta alanine metabolism | C ₇ H ₅ NO ₄ | CHEBI:16675 |
| Glycerate | 105.0 / 75.0 | Amino Acid metabolism/Gly, Ser | C ₃ H ₅ O ₄ | CHEBI:33871 |
| N-Acetyl glycine | 116.0 / 74.0 | Amino Acid metabolism | C ₄ H ₇ NO ₃ | CHEBI:40410 |
| Citraconic Acid | 129.0 / 85.0 | Amino Acid metabolism/Val, Leu, IL | C ₅ H ₆ O ₄ | CHEBI:17626 |
| D-Leucic Acid | 131.0 / 85.0 | Amino Acid metabolism/Leu | C ₆ H ₁₂ O ₃ | CHEBI:55534 |
| Glutaric Acid | 131.0 / 87.0 | Amino Acids/lys, trp, fatty acids | C ₅ H ₈ O ₄ | CHEBI:17859 |
| 2-Hydroxyglutarate | 147.0 / 129.0 | Amino-acid metabolism and Glycine/Serine/Threonine metabolism | C ₅ H ₇ O ₅ | CHEBI:132941 |
| 2-Aminoadipate | 160.0 / 116.0 | Amino acids metabolism/Lys | C ₆ H ₉ NO ₄ | CHEBI:84981 |
| PPA | 163.0 / 91.0 | Amino acids metabolism/Phe | C ₉ H ₁₃ NO | CHEBI:8104 |
| Pentothenate | 218.1 / 88.0 | Amino acids metabolism/alanine, CoA | C ₉ H ₁₆ NO ₅ | CHEBI:16454 |
| Cystathionine | 221.1 / 134.0 | Amino acids metabolism/cys | C ₇ H ₁₄ N ₂ O ₄ S | CHEBI:17755 |
| Tryptophan | 205.1 / 146.0 | Tryptophan Cycle | C ₁₁ H ₁₂ N ₂ O ₂ | CHEBI:27897 |
| L-Kynurenine | 209.1 / 94.0 | Tryptophan Cycle | C ₁₀ H ₁₂ N ₂ O ₃ | CHEBI:16946 |
| Kynurenic Acid | 188.0 / 144.0 | Tryptophan Cycle/Amino Acid metabolism | C ₁₀ H ₇ NO ₃ | CHEBI:18344 |

Continuation of Table A.1

| Name | Mass / RT | Pathway | Formula | ChEBI ID |
|-------------------------|---------------|--|---------------|-------------|
| Xanthurenic Acid | 204.1 / 160.0 | Tryptophan Cycle/Amino Acid metabolism | C10H7NO4 | CHEBI:10072 |
| Cytidine | 244.2 / 112.1 | Nucleotide/Pyrimidine metabolism | C9H13N3O5 | CHEBI:17562 |
| Uridine | 245.2 / 113.1 | Nucleotide/Pyrimidine metabolism | C9H12N2O6 | CHEBI:16704 |
| Adenosine | 268.2 / 136.1 | Nucleotide/Purine metabolism | C10H13N5O4 | CHEBI:16335 |
| Inosine | 269.2 / 137.1 | Nucleotide/Purine metabolism | C10H12N4O5 | CHEBI:17596 |
| 1-Methyladenosine | 282.2 / 150.1 | Nucleotide/Purine metabolism | C11H15N5O4 | CHEBI:16020 |
| Guanosine | 284.2 / 152.1 | Nucleotide/Purine metabolism | C10H13N5O5 | CHEBI:16750 |
| 8-Oxo-2'-deoxyguanosine | 284.2 / 168.1 | Nucleotide/Purine metabolism | C10H13N5O5 | CHEBI:40304 |
| 1-Methylguanosine | 298.2 / 166.1 | Nucleotide/Purine metabolism | C11H15N5O5 | CHEBI:19062 |
| GMP | 364.1 / 152.0 | Nucleotide/Purine metabolism | C10H14N5O8P | CHEBI:17345 |
| Uracil | 111.0 / 42.0 | Nucleotide/Pyrimidine metabolism | C4H4N2O2 | CHEBI:17568 |
| Adenine | 134.0 / 107.0 | Nucleotide/Purine metabolism | C5H5N5 | CHEBI:16708 |
| Hypoxanthine | 135.0 / 92.0 | Nucleotide | C5H4N4O | CHEBI:17368 |
| Xanthine | 151.0 / 108.0 | Nucleotide | C5H4N4O2 | CHEBI:15318 |
| Allantoin | 157.0 / 114.0 | Nucleotide Degradation | C4H6N4O3 | CHEBI:15676 |
| Urate | 167.0 / 124.0 | Nucleotide/Purine metabolism | C5H3N4O3 | CHEBI:46818 |
| Xanthosine | 283.1 / 151.0 | Nucleotide/Purine metabolism | C10H12N4O6 | CHEBI:18107 |
| DTMP | 321.1 / 79.0 | Nucleotide/pyrimidine | C10H15N2O8P | CHEBI:17013 |
| CMP | 322.0 / 97.0 | Nucleotide/Pyrimidine metabolism | C9H14N3O8P | CHEBI:17361 |
| cGMP | 344.0 / 150.0 | Nucleotide/Purine metabolism | C10H12N5O7P | CHEBI:16356 |
| AMP | 346.1 / 79.0 | Nucleotide | C10H14N5O7P | CHEBI:16027 |
| IMP | 347.1 / 79.0 | Nucleotide/Purine metabolism | C10H13N4O8P | CHEBI:17202 |
| DCDP | 386.0 / 159.0 | Nucleotide/Pyrimidine metabolism | C9H15N3O10P2 | CHEBI:28846 |
| PRPP | 389.0 / 79.0 | Nucleotide/Purine metabolism | C5H13O14P3 | CHEBI:17111 |
| ADP | 426.0 / 134.0 | Nucleotide/Purine metabolism | C10H15N5O10P2 | CHEBI:16761 |
| Choline | 104.0 / 60.0 | Vitamins | C5H14NO | CHEBI:15354 |

Continuation of Table A.1

| Name | Mass / RT | Pathway | Formula | ChEBI ID |
|----------------------|------------------|--|---------------|-------------|
| 4-Pyridoxic Acid | 182.1 / 138.0 | Vitamins/B6 | C8H9NO4 | CHEBI:17405 |
| 4-Hydroxybutyrate | 105.0 / 77.0 | Lipids/phospholipids, lig- and | C4H7O3 | CHEBI:16724 |
| Deoxycarnitine | 147.0 / 87.0 | Lipids/phospholipids, lig- and | C7H15NO2 | CHEBI:16244 |
| Glycerol-3-P | 171.0 / 79.0 | Lipids/Glycerollipid | C3H9O6P | CHEBI:15978 |
| 13-HODE | 295.1 / 195.0 | Lipids/phospholipids, lig- and | C18H32O3 | CHEBI:72639 |
| Arachidonate | 303.3 / 59.0 | Lipids/phospholipids, lig- and | C20H31O2 | CHEBI:32395 |
| 12-HETE | 319.2 / 179.0 | Lipids/phospholipids, lig- and | C20H32O3 | CHEBI:19138 |
| Acetylcarnitine | 204.1 / 85.0 | Fatty acid metabolism | C9H17NO4 | CHEBI:73024 |
| Azelaic Acid | 187.0 / 125.0 | Fatty acid metabolism | C9H16O4 | CHEBI:48131 |
| Margaric Acid | 269.1 / 251.3 | Fatty acid metabolism | C17H34O2 | CHEBI:32365 |
| Linolenic Acid | 277.1 / 259.0 | Fatty acid metabolism | C18H30O2 | CHEBI:25048 |
| Linoleic Acid | 279.1 / 261.0 | Fatty acid metabolism | C18H32O2 | CHEBI:17351 |
| Glyceraldehyde | 89.0 / 59.0 | Sugar | C3H6O3 | CHEBI:5445 |
| Inositol | 179.0 / 87.0 | Glucose/inositol metabolism | C6H12O6 | CHEBI:24848 |
| Sorbitol | 181.0 / 89.0 | Sugar | C6H14O6 | CHEBI:30911 |
| Glucuronate | 193.0 / 73.0 | Amino sugar and nucleotide sugar metabolism | C6H9O7 | CHEBI:24297 |
| Lactose | 341.0 / 59.0 | Sugar/Galactose | C12H22O11 | CHEBI:17716 |
| Sucrose | 341.0 / 59.0 (2) | Sugar | C12H22O11 | CHEBI:17992 |
| UDP-GlcNAc | 606.0 / 385.0 | Amino sugar and nucleotide sugar metabolism | C17H27N3O17P2 | CHEBI:16264 |
| Ornithine | 133.0 / 70.0 | Urea cycle | C5H12N2O2 | CHEBI:18257 |
| Citrulline | 174.0 / 131.0 | Urea Cycle | C6H13N3O3 | CHEBI:18211 |
| Benzoic Acid | 121.0 / 77.0 | Nicotinate and nicotinamide metabolism | C7H6O2 | CHEBI:30746 |
| Reduced Glutathione | 306.3 / 143.1 | Oxidative Damage | C10H17N3O6S | CHEBI:16856 |
| Oxidized Glutathione | 611.2 / 306.0 | Oxidative Damage | C10H17N3O6S | CHEBI:16856 |
| Oxalic Acid | 89.0 / 61.0 | Glyoxylate and dicarboxy- late metabolism | C2H2O4 | CHEBI:16995 |
| Acetoacetate | 101.0 / 57.0 | Ketone | C4H5O3 | CHEBI:13705 |
| Putrescine | 89.0 / 72.0 | Polyamine Metabolism | C4H12N2 | CHEBI:17148 |
| Agmatine | 131.0 / 72.0 | Polyamine Metabolism | C5H14N4 | CHEBI:17431 |
| N-Acetylputrescine | 131.0 / 114.0 | Polyamine Metabolism | C6H14N2O | CHEBI:17768 |
| Spermidine | 146.0 / 72.0 | Polyamine Metabolism | C7H19N3 | CHEBI:16610 |
| Ribose-5-P | 229.0 / 79.0 | Pentose phosphate pathway | C5H11O8P | CHEBI:52742 |

Table A.2 Significantly age-associated metabolites relationship to known metabolic changes in the literature . ↑–metabolite level increases with advanced age. ↓–metabolite level decreases with advanced age.

| Metabolite | ↑ / ↓ | Literature |
|------------------|-------|---|
| Oxaloacetate | ↓ | <i>Increased</i> with age in (Wan et al., 2017), in <i>glp-1</i> , although notably much less so (only barely significant) in <i>N2</i> and in the double mutant <i>daf-16;glp-1</i> . |
| Glutamine | ↓ | Decreased with age in (Wan et al., 2017) in <i>N2</i> , <i>glp-1</i> and the double mutant <i>daf-16;glp-1</i> . Decreased with age in (Davies et al., 2015) in <i>N2</i> but not in <i>daf-2</i> . <i>Increased</i> with age in <i>N2</i> and <i>slcf-1</i> strains in (Pontoizeau et al., 2014), but not <i>eat-2</i> . Decreased with age in (Gao et al., 2017). |
| 2-Aminoadipate | ↓ | Not previously reported. |
| Guanosine | ↓ | Reported as differing between strains in (Wan et al., 2017), but not as varying significantly with age. |
| Cystathionine | ↓ | Decreased with age in (Wan et al., 2017) in <i>N2</i> , <i>glp-1</i> and the double mutant <i>daf-16;glp-1</i> . Decreased with age in <i>daf-2</i> , but not <i>N2</i> , in (Davies et al., 2015). Also decreased with age in all strains in (Pontoizeau et al., 2014). |
| Cytidine | ↓ | Decreased with age in (Copes et al., 2015). |
| Asparagine | ↓ | Decreased with age in (Gao et al., 2017). <i>Increased</i> in <i>glp-1</i> , but no significant increase in <i>N2</i> and slightly decreased in double mutant <i>daf-16;glp-1</i> reported in (Wan et al., 2017). |
| Uridine | ↓ | Not previously reported. |
| GMP | ↓ | Not previously reported. |
| Methionine | ↓ | Decreased with age in (Copes et al., 2015) in <i>glp-4</i> . Not changing with age in (Davies et al., 2015). Decreased with age in (Gao et al., 2017). |
| Aconitate | ↓ | Not previously reported. |
| Xanthurenic Acid | ↓ | Decreased with age in (Wan et al., 2017). |
| Phenylalanine | ↓ | Decreased with age in (Copes et al., 2015) in <i>glp-4</i> , and in (Wan et al., 2017) in <i>glp-1</i> (not significant in <i>N2</i> or the double mutant <i>daf-16;glp-1</i>). Decreased with age in <i>daf-2</i> but not in <i>N2</i> in (Davies et al., 2015). Decreased with age in <i>N2</i> and <i>slcf-1</i> but not <i>eat-2</i> in (Pontoizeau et al., 2014). Decreased with age in (Gao et al., 2017). |
| iso-Leucine | ↓ | Decreased with age in <i>daf-2</i> but not <i>N2</i> in (Davies et al., 2015). Decreased with age in WT and <i>slcf-1</i> in (Pontoizeau et al., 2014). Decreased with age in (Gao et al., 2017). |
| Lysine | ↓ | Decreased with age in (Gao et al., 2017). Decreased with age in <i>glp-1</i> and in the double mutant <i>glp-1;daf-16</i> , but interestingly, not in <i>N2</i> , in (Wan et al., 2017). Decreased with age in <i>daf-2</i> but not in <i>N2</i> in (Davies et al., 2015). Decreased with age in all three strains in (Pontoizeau et al., 2014). |
| Tryptophan | ↓ | Decreased with age in (Gao et al., 2017). <i>Increased</i> with age in (Wan et al., 2017) in <i>N2</i> . |

Continuation of Table A.2 (Age-associated metabolites)

| Metabolite | ↑ / ↓ | Literature |
|--------------------------|-------|---|
| Leucine | ↓ | Decreased with age in (Wan et al., 2017) in <i>N2</i> , <i>glp-1</i> and the double mutant <i>daf-16;glp-1</i> . Decreased with age in (Gao et al., 2017). |
| Homoserine | ↓ | <i>Increased</i> with age in (Wan et al., 2017). |
| Linolenic Acid (C18:3) | ↓ | Decreasing with age in (Gao et al., 2017). |
| Threonine | ↓ | <i>Increased</i> in <i>N2</i> and <i>glp-1</i> , but decreased in the double mutant <i>daf-16;glp-1</i> in (Wan et al., 2017). Increased in <i>daf-2</i> , no change in <i>N2</i> in (Davies et al., 2015). |
| Arginine | ↓ | Decreased with age in (Gao et al., 2017). <i>Increased</i> with age in (Wan et al., 2017) in <i>N2</i> , less so in <i>glp-1</i> and not at all in the double mutant <i>glp-1;daf-16</i> . Decreased in <i>daf-2</i> , no change in <i>N2</i> in (Davies et al., 2015). Decreased in <i>N2</i> and <i>slcf-1</i> but not <i>eat-2</i> in (Pontoizeau et al., 2014). |
| Agmatine | ↓ | Not previously reported. |
| Serine | ↓ | Decreased with age in (Wan et al., 2017) in <i>N2</i> , <i>glp-1</i> and the double mutant <i>daf-16;glp-1</i> . <i>Increased</i> with age in <i>N2</i> , and only very slightly in <i>daf-2</i> , in (Davies et al., 2015). Decreased with age in (Gao et al., 2017). |
| Linoleic Acid (C18:2) | ↓ | Appears to initially increase with age, but then decrease again at D10 in (Gao et al., 2017). |
| Hypoxanthine | ↑ | <i>Decreased</i> with age in (Copes et al., 2015) in <i>glp-4</i> and <i>N2</i> (with FuDR). |
| Uracil | ↑ | Not previously reported. |
| D-Leucic Acid | ↑ | Not previously reported. |
| Putrescine | ↑ | Increasing with age in wild type, but not <i>daf-2</i> , in (Davies et al., 2015). |
| 2-Hydroxyisovaleric Acid | ↑ | Not previously reported. |
| Lactate | ↑ | No change with age in (Davies et al., 2015). <i>Decreased</i> with age in <i>N2</i> , but not other strains in (Pontoizeau et al., 2014). |
| Betaine | ↑ | No change with age in (Davies et al., 2015) or (Pontoizeau et al., 2014). |
| Choline | ↑ | No change with age in (Davies et al., 2015). Increased with age in <i>eat-2</i> but not other strains in (Pontoizeau et al., 2014). |
| Kynurenic Acid | ↑ | Not previously reported. |
| Acetylcarnitine | ↑ | Not previously reported. |
| Carnitine | ↑ | Not previously reported. |
| Succinate | ↑ | <i>Decreased</i> with age in (Wan et al., 2017) in <i>N2</i> , <i>glp-1</i> and the double mutant <i>daf-16;glp-1</i> . Also decreased with age in <i>daf-2</i> and <i>N2</i> in (Davies et al., 2015). Not significantly changed in (Pontoizeau et al., 2014). |
| Pentothenate | ↑ | Not previously reported. |
| 1/3-Methylhistidine | ↑ | Not previously reported. |
| Xanthosine | ↑ | Not previously reported. |
| 3-Hydroxybutyric Acid | ↑ | Not previously reported. |
| Allantoin | ↑ | <i>Decreased</i> with age in (Wan et al., 2017) in <i>N2</i> and double mutant <i>daf-16;glp-1</i> , only barely significant in <i>glp-1</i> . |

Continuation of Table A.2 (Age-associated metabolites)

| Metabolite | ↑ / ↓ | Literature |
|--------------------|-------|--------------------------|
| N-Acetylputrescine | ↑ | Not previously reported. |
| Cadaverine | ↑ | Not previously reported. |

Table A.3 Longevity modulating metabolites in the literature and DrugAge database.

| Metabolite | DrugAge | Other literature |
|----------------------------|---------|--|
| Pyruvate | No | Extends lifespan when supplemented, reported in (Mishur et al., 2016). |
| Lactate | Yes | |
| Oxaloacetate | Yes | |
| PEP | Yes | |
| DHAP | No | |
| D-GA3P | No | |
| Glucose | No | |
| G1P/G6P/F6P/F1P | No | |
| 3HBA | Yes | |
| Maleic Acid | No | |
| Succinate | Yes | |
| Malate | Yes | |
| Aconitate | No | |
| Glycine | Yes | |
| Alanine | Yes | |
| Cadaverine | No | |
| Serine | Yes | |
| Proline | Yes | |
| Valine | Yes | |
| Threonine | Yes | |
| Leucine | Yes | |
| iso-Leucine | Yes | |
| Asparagine | Yes | |
| Tyramine | No | |
| Glutamine | Yes | |
| Lysine | Yes | |
| Glutamic Acid | Yes | |
| Methionine | Yes | |
| Histidine | Yes | |
| Phenylalanine | No | |
| 1/3-Methylhistidine | No | |
| Arginine | Yes | |
| Tyrosine | Yes | |
| Cystine | No | |
| 2-Hydroxyisovaleric Acid | No | Extends lifespan when supplemented, reported in (Mishur et al., 2016). |
| 3-Methyl-2-Oxovaleric Acid | No | |
| Aminoisobutyrate | No | |
| Betaine | Yes | |

Continuation of Table A.3 (Longevity-modulating metabolites)

| Metabolite | DrugAge | Other literature |
|-------------------------|---------|--|
| Homoserine | No | Increased lifespan 'slightly' when supplemented in (Copes et al., 2015). |
| Creatine | No | |
| Aspartic Acid | Yes | |
| Carnitine | Yes | |
| Quinolinic Acid | No | |
| Glycerate | No | |
| N-Acetylglycine | No | |
| Citraconic Acid | No | |
| D-Leucic Acid | No | |
| Glutaric Acid | No | |
| 2-Hydroxyglutarate | Yes | |
| 2-Aminoadipate | No | |
| Pentothenate | No | |
| Cystathionine | No | |
| Tryptophan | Yes | |
| L-Kynurenine | No | |
| Kynurenic Acid | No | |
| Xanthurenic Acid | No | |
| Cytidine | No | |
| Uridine | No | |
| Adenosine | No | |
| Inosine | No | |
| 1-Methyladenosine | No | |
| Guanosine | No | |
| 8-Oxo-2'-deoxyguanosine | No | |
| 1-Methylguanosine | No | |
| GMP | No | |
| Uracil | No | |
| Adenine | No | |
| Hypoxanthine | No | |
| Xanthine | No | |
| Allantoin | Yes | |
| Urate | No | |
| Xanthosine | No | |
| CMP | No | |
| cGMP | No | |
| AMP | No | |
| IMP | No | |
| PRPP | No | |
| Choline | No | |
| 4-Pyridoxic Acid | No | |
| 4-Hydroxybutyrate | No | |
| Glycerol-3-P | No | |
| 13-HODE | No | |

Continuation of Table A.3 (Longevity-modulating metabolites)

| Metabolite | DrugAge | Other literature |
|----------------------|---------|--|
| Arachidonate | No | Extends lifespan when supplemented, reported in (O'Rourke et al., 2013). |
| 12-HETE | No | |
| Acetylcarnitine | No | Extends lifespan when supplemented, reported in (Qi et al., 2017). |
| Azelaic Acid | No | |
| Margaric Acid | No | |
| Linolenic Acid | No | |
| Linoleic Acid | No | |
| Glyceraldehyde | No | |
| Inositol | Yes | |
| Sorbitol | No | |
| Glucuronate | No | |
| Lactose | No | |
| UDP-GlcNAc | No | |
| Ornithine | Yes | |
| Citrulline | No | |
| Benzoic Acid | No | |
| Reduced Glutathione | Yes | |
| Oxidized Glutathione | No | |
| Acetoacetate | No | |
| Putrescine | Yes | |
| Agmatine | Yes | |
| N-Acetylputrescine | Yes | |
| Spermidine | Yes | |
| Ribose-5-P | No | |

Table A.4 Pathway metrics

| Pathway | No. reactions | No. blocked | % blocked |
|---|---------------|-------------|-----------|
| Acetyl-CoA conversions | 23 | 5 | 22% |
| Alanine, aspartate and glutamate metabolism | 49 | 7 | 14% |
| Amino sugar and nucleotide sugar metabolism | 32 | 9 | 28% |
| Amino sugar metabolism | 10 | 6 | 60% |
| Aminoacyl-tRNA biosynthesis | 40 | 0 | 0% |
| Arachidonic acid metabolism | 18 | 1 | 6% |
| Arginine and proline metabolism | 46 | 4 | 9% |
| Ascaroside biosynthesis | 379 | 0 | 0% |
| Ascorbate and aldarate metabolism | 9 | 0 | 0% |
| Bacterial digestion | 31 | 0 | 0% |
| Beta-alanine metabolism | 17 | 9 | 53% |
| Biomass assembly | 23 | 0 | 0% |
| Biosynthesis of unsaturated fatty acids | 15 | 0 | 0% |
| Branched-chain fatty acid synthesis | 21 | 18 | 86% |
| Butanoate metabolism | 4 | 0 | 0% |
| Carnitine shuttle | 60 | 60 | 100% |
| Chitin degradation | 6 | 5 | 83% |

Continuation of Table A.4, pathway metrics

| Pathway | No. reactions | No. blocked | % blocked |
|--|---------------|-------------|-----------|
| Chondroitin and dermatan biosynthesis | 5 | 5 | 100% |
| Chondroitin sulfate degradation | 3 | 3 | 100% |
| Collagen biosynthesis | 9 | 0 | 0% |
| Core modifications of n-glycans | 26 | 26 | 100% |
| Cyanoamino acid metabolism | 7 | 0 | 0% |
| Cysteine and methionine metabolism | 41 | 6 | 15% |
| Ether lipid metabolism | 21 | 0 | 0% |
| Exchange reactions | 284 | 74 | 26% |
| Extracellular transport | 115 | 26 | 23% |
| Fatty acid biosynthesis | 128 | 9 | 7% |
| Fatty acid metabolism | 25 | 0 | 0% |
| Fatty acid oxidation | 185 | 56 | 30% |
| Folate biosynthesis | 40 | 11 | 28% |
| Fructose and mannose metabolism | 26 | 11 | 42% |
| Galactose metabolism | 16 | 4 | 25% |
| Gamma-glutamyl cycle | 7 | 0 | 0% |
| Genetic information processing | 4 | 0 | 0% |
| Glutathione metabolism | 29 | 4 | 14% |
| Glycerolipid metabolism | 37 | 3 | 8% |
| Glycerophospholipid metabolism | 80 | 19 | 24% |
| Glycine, serine and threonine metabolism | 54 | 6 | 11% |
| Glycoaminoglycan-protein linkage region biosynthesis | 6 | 6 | 100% |
| Glycogen metabolism | 8 | 0 | 0% |
| Glycolysis and gluconeogenesis | 41 | 7 | 17% |
| Glyoxylate and dicarboxylate metabolism | 31 | 8 | 26% |
| Heparan sulfate degradation | 1 | 1 | 100% |
| Histidine metabolism | 13 | 4 | 31% |
| Histone modification | 14 | 13 | 93% |
| Inositol phosphate metabolism | 91 | 29 | 32% |
| Iron-sulfur cluster biosynthesis | 2 | 0 | 0% |
| Keratan sulfate degradation | 1 | 0 | 0% |
| Lipoic acid metabolism | 4 | 4 | 100% |
| Lysine degradation | 20 | 5 | 25% |
| Lysine metabolism | 15 | 0 | 0% |
| Methane metabolism | 3 | 1 | 33% |
| Methylglyoxal degradation | 5 | 1 | 20% |
| Miscellaneous | 41 | 22 | 54% |
| Mitochondrial fatty acid beta oxidation | 18 | 17 | 94% |
| Mitochondrial transport | 114 | 15 | 13% |
| Molybdenum cofactor biosynthesis | 14 | 5 | 36% |
| N-acetylglucosamine degradation | 4 | 1 | 25% |
| N-glycan biosynthesis | 82 | 44 | 54% |
| N-glycan degradation | 6 | 6 | 100% |
| Needed glycerolipid metabolism | 3 | 0 | 0% |
| Nicotinamide adenine dinucleotide metabolism | 23 | 10 | 43% |
| Nicotinate and nicotinamide metabolism | 11 | 1 | 9% |

Continuation of Table A.4, pathway metrics

| Pathway | No. reactions | No. blocked | % blocked |
|---|---------------|-------------|-----------|
| Nuclear transport | 46 | 13 | 28% |
| Nucleotide interconversion | 16 | 11 | 69% |
| O-glycan synthesis | 5 | 4 | 80% |
| One carbon pool by folate | 9 | 0 | 0% |
| Oxidative phosphorylation | 18 | 1 | 6% |
| Pantothenate and coenzyme A biosynthesis | 11 | 0 | 0% |
| Pentose and glucuronate interconversions | 10 | 1 | 10% |
| Pentose phosphate pathway | 32 | 14 | 44% |
| Peroxisomal fatty acid beta oxidation | 25 | 5 | 20% |
| Phenylalanine degradation | 10 | 2 | 20% |
| Phenylalanine metabolism | 6 | 0 | 0% |
| Phosphatidylinositol phosphate metabolism | 19 | 19 | 100% |
| Porphyrin and chlorophyll metabolism | 9 | 2 | 22% |
| Primary bile acid biosynthesis | 28 | 28 | 100% |
| Propanoate metabolism | 14 | 2 | 14% |
| Protein hydroxylation | 10 | 0 | 0% |
| Protein methylation | 3 | 3 | 100% |
| Protein modification | 9 | 5 | 56% |
| Purine metabolism | 152 | 16 | 11% |
| Pyrimidine metabolism | 120 | 11 | 9% |
| Pyruvate metabolism | 13 | 0 | 0% |
| Retinol metabolism | 19 | 17 | 89% |
| Riboflavin metabolism | 11 | 0 | 0% |
| Selenocompound metabolism | 27 | 12 | 44% |
| Sphingolipid metabolism | 35 | 5 | 14% |
| Starch and sucrose metabolism | 17 | 1 | 6% |
| Steroid biosynthesis | 13 | 13 | 100% |
| Steroid degradation | 1 | 1 | 100% |
| Steroid hormone biosynthesis | 11 | 11 | 100% |
| Straight-chain fatty acid synthesis | 26 | 26 | 100% |
| Sucrose degradation | 8 | 3 | 38% |
| Sulfur metabolism | 15 | 1 | 7% |
| Taurine and hypotaurine metabolism | 7 | 0 | 0% |
| TCA cycle | 31 | 6 | 19% |
| Terpenoid backbone biosynthesis | 37 | 4 | 11% |
| Thiamine metabolism | 3 | 0 | 0% |
| Thioredoxin pathway | 3 | 0 | 0% |
| Transport | 293 | 81 | 28% |
| Trehalose degradation | 3 | 1 | 33% |
| Triterpenoid biosynthesis | 3 | 3 | 100% |
| Tryptophan metabolism | 42 | 1 | 2% |
| Tyrosine biosynthesis | 2 | 1 | 50% |
| Tyrosine metabolism | 32 | 9 | 28% |
| Ubiquinol-9 biosynthesis | 18 | 2 | 11% |
| Ubiquinone and other terpenoid-quinone biosynthesis | 11 | 0 | 0% |
| Valine, leucine and isoleucine degradation | 68 | 18 | 26% |

Continuation of Table A.4, pathway metrics

| Pathway | No. reactions | No. blocked | % blocked |
|------------------------|---------------|-------------|-----------|
| Vitamin B12 metabolism | 8 | 0 | 0% |
| Vitamin B6 metabolism | 13 | 4 | 31% |

Table A.5 Essential genes in WormJam model

| Gene ID (WormBase) | Gene name | Description |
|--------------------|------------|--|
| WBGene00000197 | aars-2 | ortholog of human alanyl-tRNA synthetase |
| WBGene00004679 | rars-1 | ortholog of human arginyl-tRNA synthetase |
| WBGene00003815 | nars-1 | ortholog of human asparaginyl-tRNA synthetase |
| WBGene00000800 | cars-1 | predicted cysteine-tRNA ligase activity |
| WBGene00001744 | gars-1 | ortholog of human glycyl-tRNA synthetase |
| WBGene00002152 | iars-1 | ortholog of human isoleucyl-tRNA synthetase |
| WBGene00013678 | Y105E8A.20 | ortholog of human methionyl-tRNA synthetase, mitochondrial |
| WBGene00001497 | fars-1 | ortholog of human phenylalanyl-tRNA synthetase, alpha subunit |
| WBGene00005663 | sars-1 | ortholog of human seryl-tRNA synthetase |
| WBGene00006617 | tars-1 | ortholog of human threonyl-tRNA synthetase |
| WBGene00006945 | wars-1 | ortholog of human tryptophanyl-tRNA synthetase |
| WBGene00001094 | dars-1 | ortholog of human aspartyl-tRNA synthetase |
| WBGene00005662 | sars-2 | ortholog of human seryl-tRNA synthetase 2, mitochondrial |
| WBGene00000196 | aars-1 | ortholog of human alanyl-tRNA synthetase |
| WBGene00004680 | rars-2 | ortholog of human arginyl-tRNA synthetase 2, mitochondrial |
| WBGene00013447 | nars-2 | ortholog of human asparaginyl-tRNA synthetase 2, mitochondrial |
| WBGene00002153 | iars-2 | ortholog of human isoleucyl-tRNA synthetase 2, mitochondrial |
| WBGene00003074 | lars-2 | ortholog of human leucyl-tRNA synthetase 2, mitochondrial |
| WBGene00013361 | fars-2 | ortholog of human phenylalanyl-tRNA synthetase 2, mitochondrial |
| WBGene00004190 | pars-2 | ortholog of human prolyl-tRNA synthetase 2, mitochondrial |
| WBGene00006946 | prx-10 | ortholog of human tryptophanyl-tRNA synthetase 2, mitochondrial |
| WBGene00006935 | vars-1 | predicted to have aminoacyl-tRNA editing activity |
| WBGene00016524 | C39B5.6 | ortholog of human glutamyl-tRNA amidotransferase subunit B |
| WBGene00013433 | Y66D12A.7 | ortholog of human glutamyl-tRNA amidotransferase subunit C |
| WBGene00021508 | Y41D4A.6 | ortholog of human glutaminyl-tRNA amidotransferase subunit A |
| WBGene00001095 | dars-2 | ortholog of human aspartyl-tRNA synthetase 2, mitochondrial |
| WBGene00020696 | T22F3.3 | ortholog of human glycogen phosphorylase |
| WBGene00002891 | let-767 | ortholog of human hydroxysteroid 17-beta dehydrogenase 12 |
| WBGene00020517 | hpo-8 | ortholog of human HACD1 and HACD2 (3-hydroxyacyl-CoA dehydratases) |
| WBGene00000198 | art-1 | ortholog of human TECR and TECRL (enoyl-CoA reductases) |
| WBGene00001243 | elo-5 | predicted acyltransferase |
| WBGene00009342 | fasn-1 | ortholog of human FASN (fatty acid synthase) |
| WBGene00000254 | bli-4 | ortholog of human PCSK5 (proprotein convertase) |
| WBGene00000253 | bli-3 | ortholog of human DUOX1 and DUOX2 (dual oxidases) |
| WBGene00001508 | fut-8 | ortholog of human FUT8 (fucosyltransferase) |
| WBGene00019322 | ahcy-1 | ortholog of human ACHY (adenosylhomocysteinase) |

Continuation of Table A.5, essential genes

| Gene ID (WormBase) | Gene name | Description |
|--------------------|-----------|--|
| WBGene00013024 | cept-2 | ortholog of human CEPT1 and CHPT1 (phosphotransferases) |
| WBGene00009057 | cept-1 | ortholog of human CEPT1 and CHPT1 (phosphotransferases) |
| WBGene00012936 | pola-1 | ortholog of human POLA1 (DNA polymerase alpha 1 catalytic subunit) |
| WBGene00009368 | pole-1 | ortholog of human POLE (DNA polymerase epsilon catalytic subunit) |
| WBGene00017237 | pole-2 | ortholog of human POLE2 (DNA polymerase epsilon 2, accessory subunit) |
| WBGene00008645 | F10C2.4 | ortholog of human POLD1 (DNA polymerase delta 1, catalytic subunit) |
| WBGene00019380 | K04C2.2 | predicted to have DNA binding and polymerase activity |
| WBGene00013258 | polg-1 | ortholog of human POLG (DNA polymerase gamma, catalytic subunit) |
| WBGene00013150 | Y53F4B.3 | ortholog of human POLE4 (DNA polymerase epsilon 4, accessory subunit) |
| WBGene00017696 | polk-1 | ortholog of human POLK (DNA polymerase kappa) |
| WBGene00020823 | T26A5.8 | predicted to have protein heterodimerization activity |
| WBGene00008722 | F12F6.7 | ortholog of human POLD2 (DNA polymerase delta 2, accessory subunit) |
| WBGene00004181 | pri-2 | ortholog of human PRIM2 (DNA primase) |
| WBGene00014115 | gld-4 | ortholog of human TENT4A and TENT4B (terminal nucleotidyltransferases 4A and 4B) |
| WBGene00008781 | rpoa-2 | ortholog of human POLR1B (RNA polymerase I subunit B) |
| WBGene00004411 | rpc-1 | ortholog of human POLR3A (RNA polymerase III subunit A) |
| WBGene00017300 | rpc-2 | ortholog of human POLR3B (RNA polymerase III subunit B) |
| WBGene00010408 | mboa-2 | ortholog of human DGAT1 (diacylglycerol O-acyltransferase) |
| WBGene00012911 | acl-7 | ortholog of human GNPAT (O-acyltransferase) |
| WBGene00016384 | cdgs-1 | ortholog of human CDS2 (phosphatidate cytidylyltransferase) |
| WBGene00021677 | pgs-1 | ortholog of human PGS1 (CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase) |
| WBGene00017763 | crls-1 | ortholog of human CRLS1 (cardiolipin synthase) |
| WBGene00022781 | pmt-1 | exhibits phosphoethanolamine N-methyltransferase activity |
| WBGene00018811 | pmt-2 | exhibits phosphatidyl-N-dimethylethanolamine N-methyltransferase activity and phosphatidyl-N-methylethanolamine N-methyltransferase activity |
| WBGene00013920 | pssy-1 | ortholog of human PTDSS1 (phosphatidylserine synthase 1) |
| WBGene00012897 | pisy-1 | ortholog of human CDIPT (CDP-diacylglycerol-inositol 3-phosphatidyltransferase) |
| WBGene00012148 | inos-1 | ortholog of human ISYNA1 (inositol-3-phosphate synthase) |
| WBGene00015021 | B0205.6 | ortholog of human NFS1 (cysteine desulfurase) |
| WBGene00012885 | Y45F10D.4 | an ortholog of human ISCU (iron ion binding and iron-sulfur cluster binding) |
| WBGene00002010 | hsp-6 | ortholog of human HSPA9 (heat shock protein family A (Hsp70) member 9) |
| WBGene00020437 | stt-3 | ortholog of human STT3B (catalytic subunit of the oligosaccharyltransferase complex) |

Continuation of Table A.5, essential genes

| Gene ID (WormBase) | Gene name | Description |
|--------------------|-----------|--|
| WBGene00015162 | algn-11 | ortholog of human ALG11 (alpha-1,2-mannosyltransferase) |
| WBGene00020820 | algn-1 | ortholog of human ALG1 (chitobiosyldiphosphodolichol beta-mannosyltransferase) |
| WBGene00008775 | mogs-1 | ortholog of human MOGS (mannosyl-oligosaccharide glucosidase) |
| WBGene00019276 | H43I07.3 | ortholog of human ALG5 (dolichyl-phosphate beta-glucosyltransferase) |
| WBGene00001645 | gly-20 | ortholog of human MGAT2 (mannosyl (alpha-1,6-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase) |
| WBGene00010720 | K09E4.2 | ortholog of human ALG3 (alpha-1,3- mannosyltransferase) |
| WBGene00022629 | ZC513.5 | ortholog of human ALG12 (alpha-1,6-mannosyltransferase) |
| WBGene00007043 | tag-179 | ortholog of human ALG10 (alpha-1,2-glucosyltransferase) |
| WBGene00001338 | ears-2 | ortholog of human EARS2 (glutamyl-tRNA synthetase 2, mitochondrial) |
| WBGene00014054 | dbt-1 | ortholog of human DBT (dihydrolipoyllysine-residue (2-methylpropanoyl)transferase) |
| WBGene00002497 | let-268 | ortholog of human PLOD1, PLOD2, and PLOD3 (procollagen-lysines, UDP-glucose:glycoprotein glucosyltransferases) |
| WBGene00011409 | T04A8.7 | ortholog of human GBE1 (1,4-alpha-glucan branching enzyme 1) |
| WBGene00001793 | gsy-1 | ortholog of human GYS1 and GYS2 (glycogen synthases 1 and 2) |
| WBGene00011050 | agl-1 | ortholog of human AGL (amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase) |
| WBGene00011058 | fdps-1 | ortholog of human FDPS (farnesyl pyrophosphate synthase) |
| WBGene00019460 | idi-1 | ortholog of human IDI1 and IDI2 (isopentenyl-diphosphate delta-isomerases) |
| WBGene00021534 | mvk-1 | ortholog of human MVK (mevalonate kinase) |
| WBGene00009335 | F32D8.13 | ortholog of human PMVK (phosphomevalonate kinase) |
| WBGene00006465 | fntb-1 | ortholog of human FNTB (farnesyltransferase, beta subunit) |
| WBGene00019823 | fnta-1 | ortholog of human FNTA (farnesyltransferase, CAAX box, alpha subunit) |
| WBGene00001149 | bcat-1 | ortholog of human BCAT1 and BCAT2 (branched chain amino-transferases) |
| WBGene00016020 | sptl-1 | ortholog of human SPTLC1 (serine C-palmitoyltransferase) |

Table A.6 Mapping of metabolomics measured metabolites to WormJam model metabolites

| Metabolomics measured metabolite | Model metabolite |
|----------------------------------|------------------|
| Pyruvate (87.0 / 43.0) | pyr_m |
| Lactate (89.0 / 43.0) | lac_L_c |
| Oxaloacetate (131.0 / 113.0) | oaa_m |
| PEP (167.0 / 79.0) | pep_c |
| DHAP (167.0 / 79.0 (2)) | dhap_c |
| D-GA3P (167.0 / 97.0) | |
| Glucose (179.0 / 89.0) | glc_aD_c |
| G1P/G6P/F6P/F1P (259.0 / 97.0) | g6p_A_c |
| 3HBA (103.0 / 59.0) | |

Continuation of Table A.6, model metabolite mapping

| Metabolomics measured metabolite | Model metabolite |
|--|------------------|
| Maleic Acid (115.0 / 71.0 (2)) | |
| Succinate (117.0 / 73.0) | succ_m |
| Malate (133.0 / 115.0) | mal_L_m |
| Aconitate (173.0 / 85.0) | acon_C_m |
| Glycine (76.0 / 30.1) | gly_c |
| Alanine (90.0 / 44.0 (2)) | ala_L_c |
| Cadaverine (103.0 / 86.0) | |
| Serine (106.0 / 60.0 (2)) | ser_L_c |
| Proline (116.0 / 70.0) | pro_L_c |
| Valine (118.0 / 72.0) | val_L_c |
| Threonine (120.0 / 74.0 (2)) | thr_L_c |
| Leucine (132.0 / 86.0) | leu_L_c |
| iso-Leucine (132.0 / 86.0 (2)) | ile_L_c |
| Asparagine (133.0 / 74.0) | asn_L_c |
| Tyramine (138.0 / 121.0) | tym_c |
| Glutamine (147.0 / 84.0) | gln_L_c |
| Lysine (147.0 / 84.0 (2)) | lys_L_c |
| Glutamic Acid (148.0 / 84.0) | glu_L_c |
| Methionine (150.0 / 61.0) | met_L_c |
| Histidine (156.0 / 110.0) | his_L_c |
| Phenylalanine (166.0 / 120.0) | phe_L_c |
| 1/3-Methylhistidine (170.0 / 96.0) | |
| Arginine (175.0 / 70.0) | arg_L_c |
| Tyrosine (182.1 / 136.0) | tyr_L_c |
| Cystine (241.1 / 120.0) | |
| 2-Hydroxyisovaleric Acid (117.0 / 71.0) | |
| 3-Methyl-2-Oxovaleric Acid (129.0 / 101.0) | |
| Aminoisobutyrate (104.0 / 86.0) | |
| Betaine (118.0 / 58.0) | glyb_c |
| Homoserine (120.0 / 74.0) | |
| Creatine (132.0 / 90.0) | |
| Aspartic Acid (134.0 / 74.0) | asp_L_c |
| Carnitine (162.0 / 85.0) | crn_c |
| Quinolinic Acid (168.0 / 150.0) | quln_c |
| Glycerate (105.0 / 75.0) | glyc_R_c |
| N-Acetylglycine (116.0 / 74.0) | |
| Citraconic Acid (129.0 / 85.0) | |
| D-Leucic Acid (131.0 / 85.0) | |
| Glutaric Acid (131.0 / 87.0) | |
| 2-Hydroxyglutarate (147.0 / 129.0) | 2hglut_c |
| 2-Aminoadipate (160.0 / 116.0) | L2aadp_c |
| Pentothenate (218.1 / 88.0) | pnto_R_c |
| Cystathionine (221.1 / 134.0) | cyst_L_c |
| Tryptophan (205.1 / 146.0) | trp_L_c |
| L-Kynurenine (209.1 / 94.0) | Lkynr_c |
| Kynurenic Acid (188.0 / 144.0) | kynate_c |

Continuation of Table A.6, model metabolite mapping

| Metabolomics measured metabolite | Model metabolite |
|---|------------------|
| Xanthurenic Acid (204.1 / 160.0) | xanate_c |
| Cytidine (244.2 / 112.1) | cytd_c |
| Uridine (245.2 / 113.1) | uri_c |
| Adenosine (268.2 / 136.1) | adn_c |
| Inosine (269.2 / 137.1) | ins_c |
| 1-Methyladenosine (282.2 / 150.1) | |
| Guanosine (284.2 / 152.1) | gsn_c |
| 8-Oxo-2'-deoxyguanosine (284.2 / 168.1) | |
| GMP (364.1 / 152.0) | gmp_c |
| Uracil (111.0 / 42.0) | ura_c |
| Adenine (134.0 / 107.0) | ade_c |
| Hypoxanthine (135.0 / 92.0) | hxan_c |
| Xanthine (151.0 / 108.0) | xan_c |
| Allantoin (157.0 / 114.0) | |
| Urate (167.0 / 124.0) | urate_c |
| Xanthosine (283.1 / 151.0) | xtsn_c |
| CMP (322.0 / 97.0) | cmp_c |
| cGMP (344.0 / 150.0) | 35cgmp_c |
| AMP (346.1 / 79.0) | amp_c |
| IMP (347.1 / 79.0) | imp_c |
| PRPP (389.0 / 79.0) | prpp_c |
| Choline (104.0 / 60.0) | chol_c |
| 4-Pyridoxic Acid (182.1 / 138.0) | |
| 4-Hydroxybutyrate (105.0 / 77.0) | |
| Glycerol-3-P (171.0 / 79.0) | glyc3p_c |
| 13-HODE (295.1 / 195.0) | |
| Arachidonate (303.3 / 59.0) | arachd_c |
| 12-HETE (319.2 / 179.0) | |
| Acetylcarnitine (204.1 / 85.0) | acrn_c |
| Azelaic Acid (187.0 / 125.0) | |
| Margaric Acid (269.1 / 251.3) | |
| Linolenic Acid (277.1 / 259.0) | |
| Linoleic Acid (279.1 / 261.0) | lnlc_c |
| Glyceraldehyde (89.0 / 59.0) | glyald_c |
| Inositol (179.0 / 87.0) | inost_c |
| Sorbitol (181.0 / 89.0) | sorbitol_c |
| Glucuronate (193.0 / 73.0) | glcur_c |
| Lactose (341.0 / 59.0) | lcts_c |
| UDP-GlcNAc (606.0 / 385.0) | uacgam_c |
| Ornithine (133.0 / 70.0) | orn_c |
| Citrulline (174.0 / 131.0) | citr_L_c |
| Benzoic acid (121.0 / 77.0) | |
| Reduced Glutathione (306.3 / 143.1) | gthrd_c |
| Oxidized Glutathione (611.2 / 306.0) | gthox_c |
| Acetoacetate (101.0 / 57.0) | acac_c |
| Putrescine (89.0 / 72.0) | ptrc_c |

Continuation of Table A.6, model metabolite mapping

| Metabolomics measured metabolite | Model metabolite |
|------------------------------------|----------------------------|
| Agmatine (131.0 / 72.0) | aprut_c spmd_c r5p_c |
| N-Acetylputrescine (131.0 / 114.0) | |
| Spermidine (146.0 / 72.0) | |
| Ribose-5-P (229.0 / 79.0) | |

Table A.7 Predicted network-based longevity modulation scores for metabolites

| Metabolite | Edges (<i>n</i>) | Centrality | Sum (d^L) | Score (d^L) |
|--------------------------|--------------------|------------|---------------|-----------------|
| Maleic Acid | 8 | 0.7555 | 3.0833 | 0.2588 |
| Aconitate | 26 | 0.7619 | 21.1180 | 0.5959 |
| G1P/G6P/F6P/F1P | 9 | 0.7473 | 3.3893 | 0.2533 |
| DHAP | 6 | 0.7107 | 3.9212 | 0.3981 |
| Cadaverine | 7 | 0.6797 | 4.6866 | 0.3982 |
| Phenylalanine | 22 | 0.7859 | 18.5575 | 0.6341 |
| Tyramine | 12 | 0.7577 | 4.4997 | 0.2623 |
| Quinolinic Acid | 13 | 0.7443 | 4.6923 | 0.2495 |
| Glycerate | 10 | 0.6839 | 7.6629 | 0.4764 |
| 2-Aminoadipate | 11 | 0.7502 | 8.6400 | 0.5401 |
| Citraconic Acid | 12 | 0.7605 | 5.1569 | 0.3017 |
| Cystathionine | 8 | 0.7177 | 5.5840 | 0.4453 |
| Kynurenic Acid | 1 | 0.6879 | 0.3210 | 0.1104 |
| Pentothenate | 22 | 0.7606 | 17.8897 | 0.5916 |
| Uridine | 7 | 0.7510 | 4.0999 | 0.3849 |
| Inosine | 5 | 0.6975 | 1.6024 | 0.1862 |
| D-GA3P | 1 | 0.9071 | 0 | 0 |
| GMP | 4 | 0.6672 | 2.8657 | 0.38245 |
| 1-Methyladenosine | 3 | 0.6674 | 1.5810 | 0.2638 |
| AMP | 3 | 0.7623 | 0.3991 | 0.0761 |
| Xanthurenic Acid | 22 | 0.7384 | 13.2443 | 0.4252 |
| Guanosine | 13 | 0.7485 | 8.8048 | 0.4707 |
| Adenine | 6 | 0.7104 | 3.8884 | 0.3946 |
| Choline | 25 | 0.7389 | 20.1090 | 0.5715 |
| 4-Hydroxybutyrate | 1 | 0.7086 | 0.1396 | 0.0494 |
| Azelaic Acid | 6 | 0.7120 | 0.9302 | 0.0946 |
| 12-HETE | 4 | 0.6981 | 1.4280 | 0.1993 |
| Acetylcarnitine | 5 | 0.6934 | 4.2668 | 0.4931 |
| cGMP | 11 | 0.7185 | 6.2664 | 0.3752 |
| N-Acetylglycine | 6 | 0.7428 | 1.8287 | 0.1940 |
| Sorbitol | 7 | 0.7091 | 2.8821 | 0.2554 |
| Margaric Acid | 4 | 0.6849 | 0.4280 | 0.0586 |
| Pyruvate | 1 | 0.6850 | 0.2829 | 0.0969 |
| Xanthosine | 6 | 0.7061 | 4.6434 | 0.4684 |
| 2-Hydroxyisovaleric Acid | 1 | 0.6683 | 0.3765 | 0.1258 |
| PRPP | 2 | 0.6563 | 1.3210 | 0.2890 |
| Glucuronate | 3 | 0.7297 | 1.3046 | 0.2379 |
| Uracil | 6 | 0.7035 | 6 | 0.6030 |

Continuation of Table A.7

| Metabolite | Edges (<i>n</i>) | Centrality | Sum (d^{L_i}) | Score (d^L) |
|--------------------|--------------------|------------|-------------------|-----------------|
| Xanthine | 2 | 0.7325 | 2 | 0.4883 |
| Urate | 1 | 0.9070 | 0 | 0 |
| IMP | 2 | 0.7901 | 0.5001 | 0.1317 |
| Glycerol-3-P | 7 | 0.6949 | 5.6090 | 0.4872 |
| 13-HODE | 1 | 0.6764 | 1 | 0.3382 |
| Arachidonate | 4 | 0.7126 | 2.8042 | 0.3996 |
| Linoleic Acid | 1 | 0.7345 | 1 | 0.3672 |
| Lactose | 2 | 0.6964 | 1.2993 | 0.3016 |
| UDP-GlcNAc | 8 | 0.7544 | 3.0666 | 0.2570 |
| Citrulline | 1 | 0.6788 | 0 | 0 |
| Benzoic Acid | 11 | 0.7760 | 4.1968 | 0.2714 |
| Acetoacetate | 4 | 0.7073 | 2 | 0.2829 |
| N-Acetylputrescine | 16 | 0.7112 | 11.6922 | 0.4891 |
| Ribose-5-P | 2 | 0.6941 | 0.2986 | 0.0691 |

Supplementary Figures

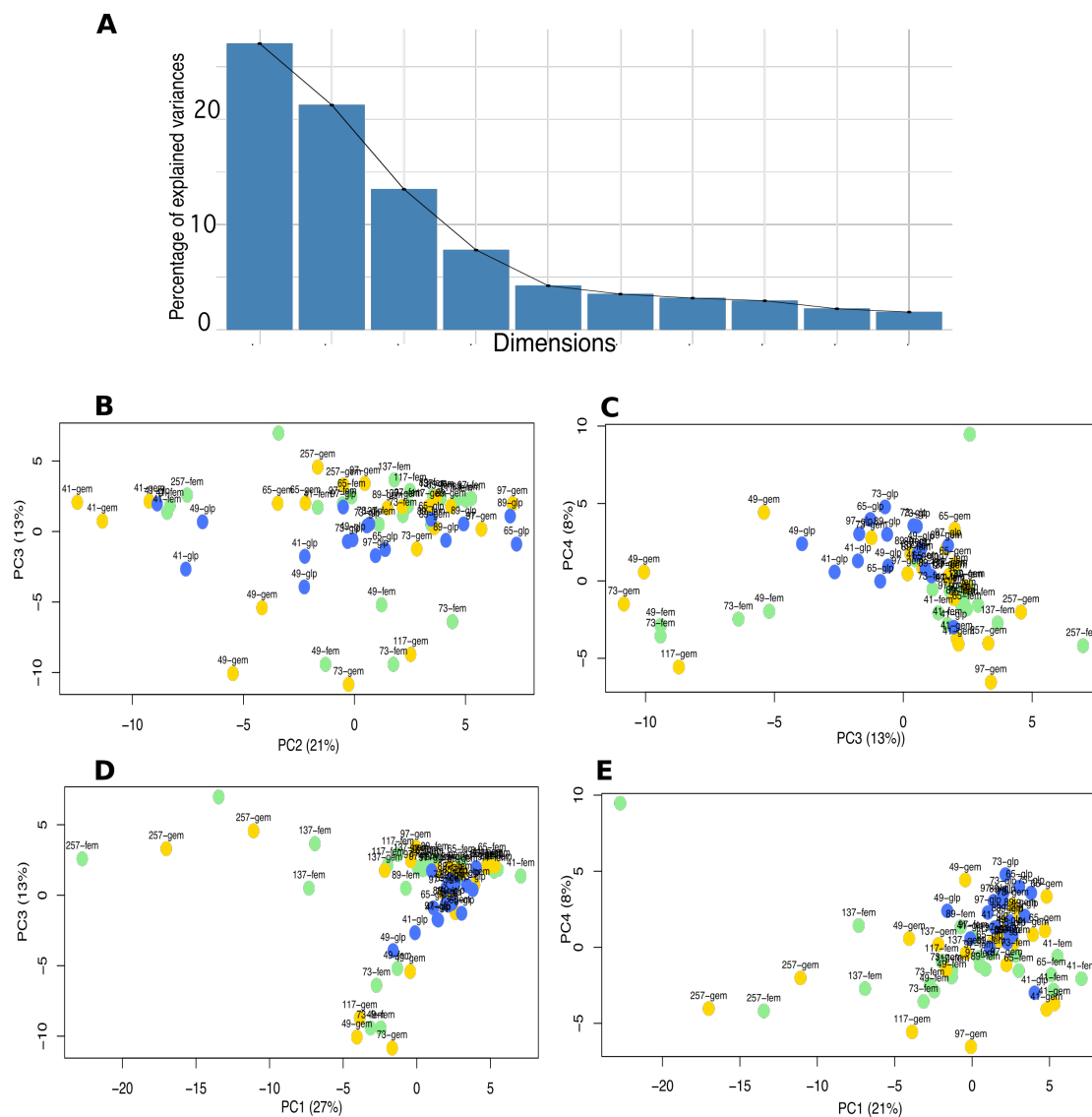


Figure A.1 Additional PCA plots for the third and fourth PC

This Figure shows (A) a breakdown of the percentages of variability explained by the first several dimensions in the PCA; (B) PC2 and PC3 (C) PC3 and PC4; (D) PC1 and PC3; (E) PC1 and PC4. All PC plots are coloured by strain – blue=GLP, gold=GEM, green=FEM.

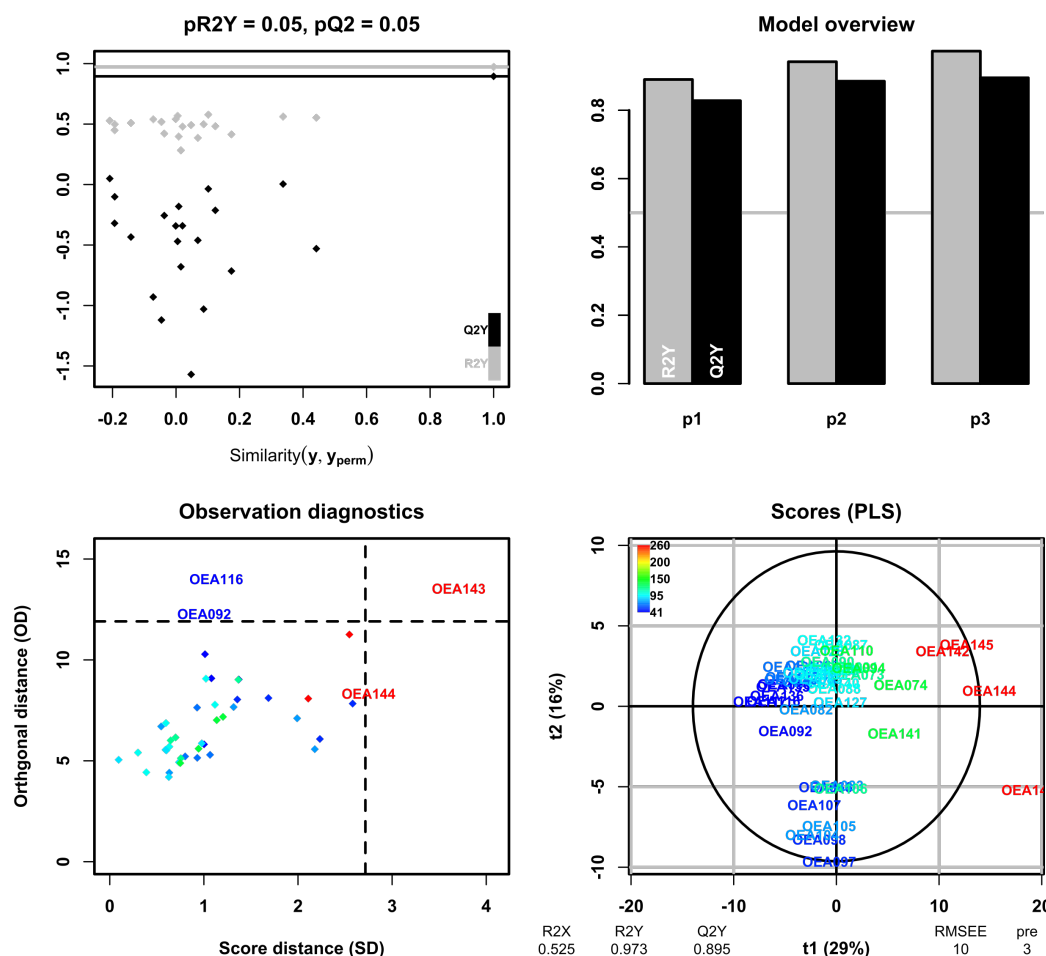


Figure A.2 The PLS-DA model of the ageing metabolome with hour of collection as response

The PLS-DA model overview, diagnostics and scores plots for the PLS-DA analysis computed by the 'ropls' library for the metabolomics dataset using hour of sample collection as the response variable are illustrated in this Figure. The summary statistics show that the model provides good predictions (Q2 is close to 1) but with the proviso that the underlying data are noisy (RMSEE is large). R2X (respectively R2Y): is the cumulative percentage of predictor (respectively response) variance explained by the full model. Q2 is the cumulative predictive performance of the model estimated by cross-validation. Pre is the number of components required to achieve predictive performance. Top left panel: significance diagnostic: the R2Y and Q2Y of the model are compared with the corresponding values obtained after random permutation of the y response; Top right panel: inertia barplot: the graphic here suggests that 3 components may be sufficient to capture most of the inertia; Bottom left panel: outlier diagnostics, suggesting that the labelled samples might be outliers (very young and old, reflecting the relatively homogenous 'middle age' with younger and older time points further removed from the remainder); Bottom right panel: x-score plot: the number of components and the cumulative R2X, R2Y and Q2Y are indicated below the plot. In the scores plot, the ellipse corresponds to 95% of the multivariate normal distribution with the samples covariance. Legends 'OEAXxx' (where x is numeric) indicate sample names. Colour represents age of sample - blue young, red old.

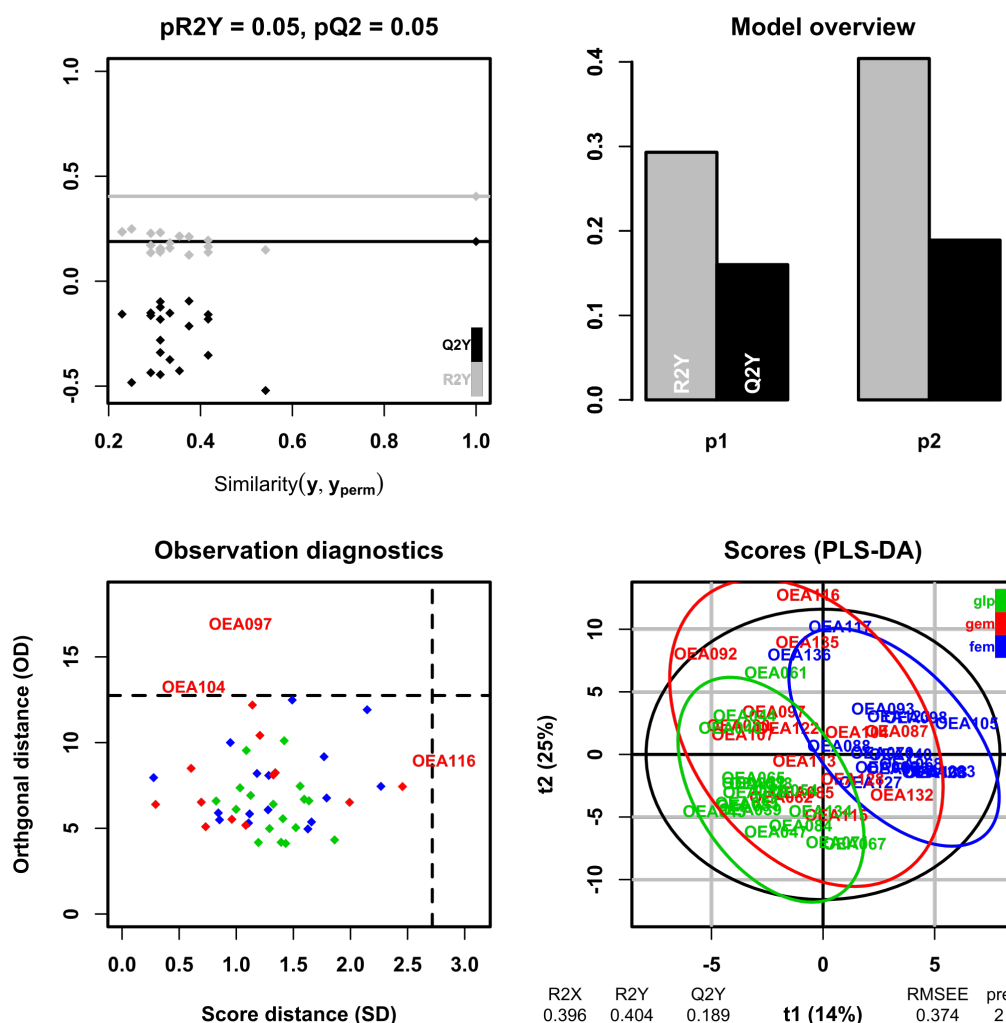


Figure A.3 The PLS-DA model of the early ageing metabolome with strain as response

The PLS-DA model overview, diagnostics and scores plots for the PLS-DA analysis computed by the 'ropls' library when strain is used as the response variable in the early ageing metabolomics dataset are illustrated in this Figure. The summary statistics show that the model is *not* robustly predictive (Q2 is close to 0) and the underlying data are noisy (RMSEE is large). R2X (respectively R2Y): is the cumulative percentage of predictor (respectively response) variance explained by the full model. Q2 is the cumulative predictive performance of the model estimated by cross-validation. Top left panel: significance diagnostic: the R2Y and Q2Y of the model are compared with the corresponding values obtained after random permutation of the y response; Top right panel: inertia barplot: the graphic here suggests that 3 components may be sufficient to capture most of the inertia; Bottom left panel: outlier diagnostics, suggesting that the labelled samples (all three from the GEM strain) might be outliers; Bottom right panel: x-score plot: the number of components and the cumulative R2X, R2Y and Q2Y are indicated below the plot. In the scores plot, the ellipses correspond to 95% of the multivariate normal distribution with the samples covariance. Legends 'OEAXxx' (where x is numeric) indicate sample names. Colour represents strain - blue for FEM, red for GEM, green for GLP.

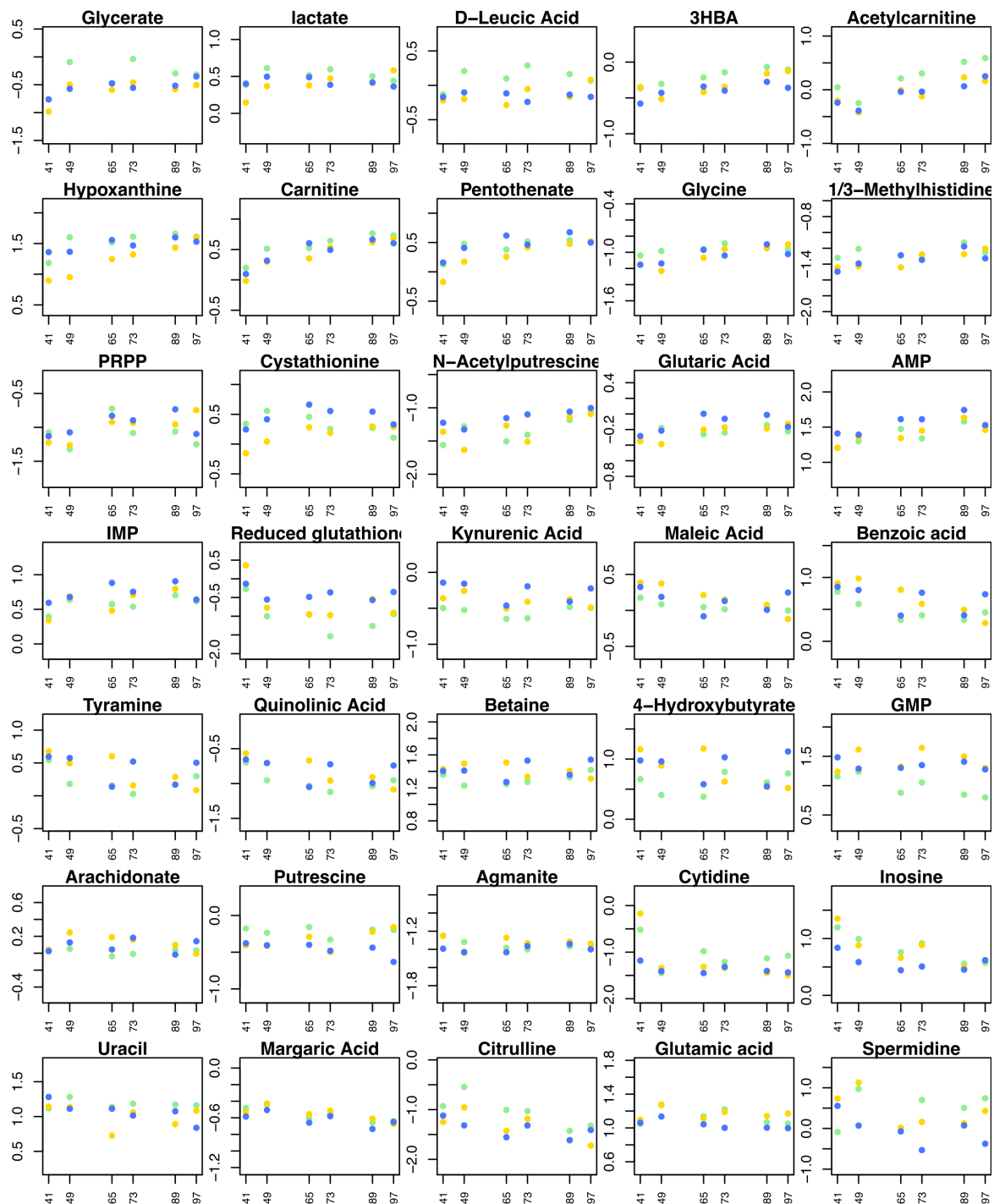


Figure A.4 Individual level plots of metabolites that differ by strain

This Figure shows the metabolites that differ by strain during early ageing. Each plot shows the normalised relative concentration levels for a given metabolite and strain at each of the time points in the early ageing dataset. Strains are indicated by colour: blue for GLP, green for FEM and gold for GEM.

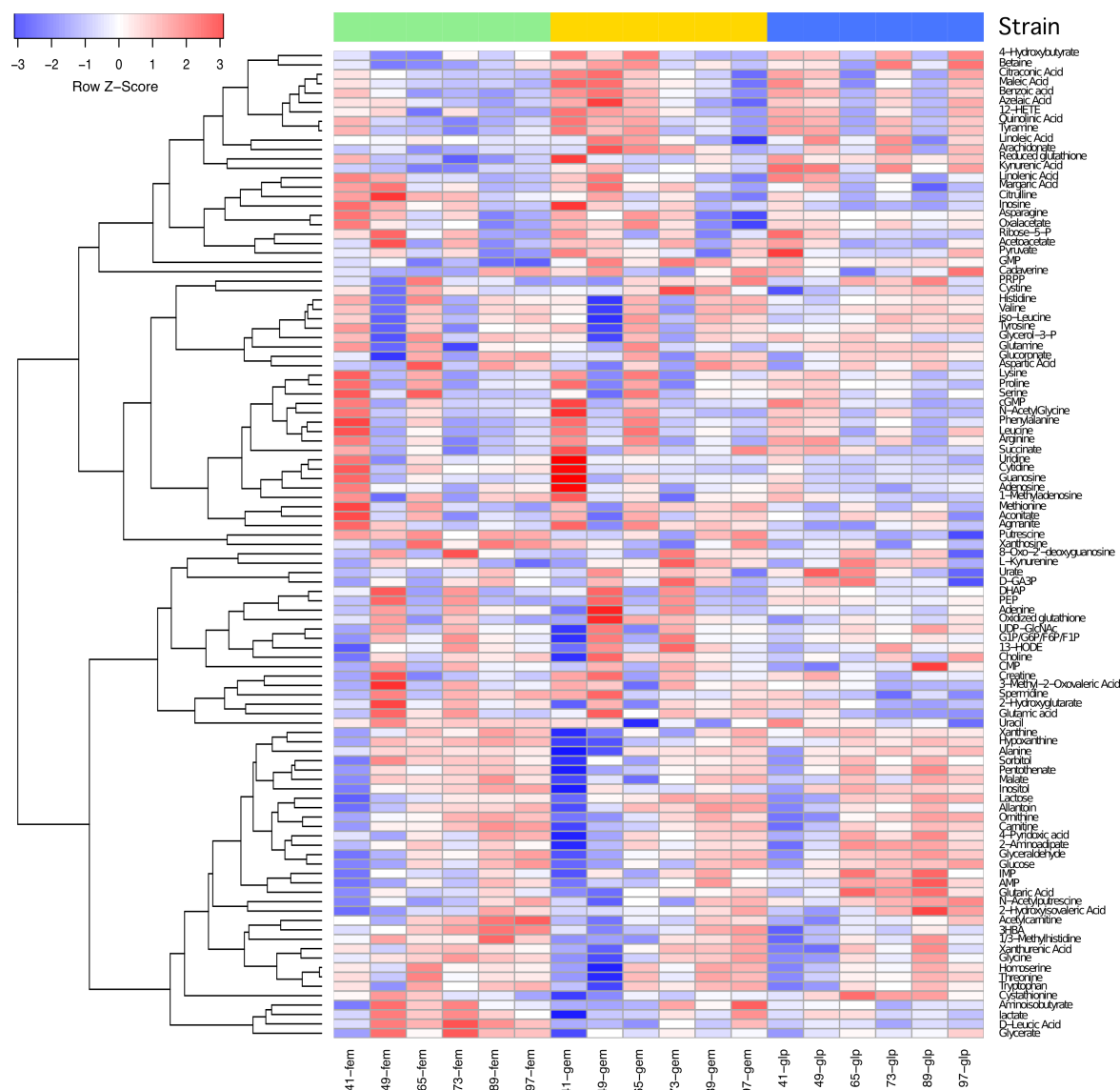


Figure A.5 Clustered heat map of full early ageing dataset

This Figure shows the full dataset of early ageing metabolomics relative concentration values as a heat map. Rows are clustered algorithmically. Columns are ordered by strain and then by time point. Strains are indicated by column top bar colour: blue for GLP, green for FEM and gold for GEM.

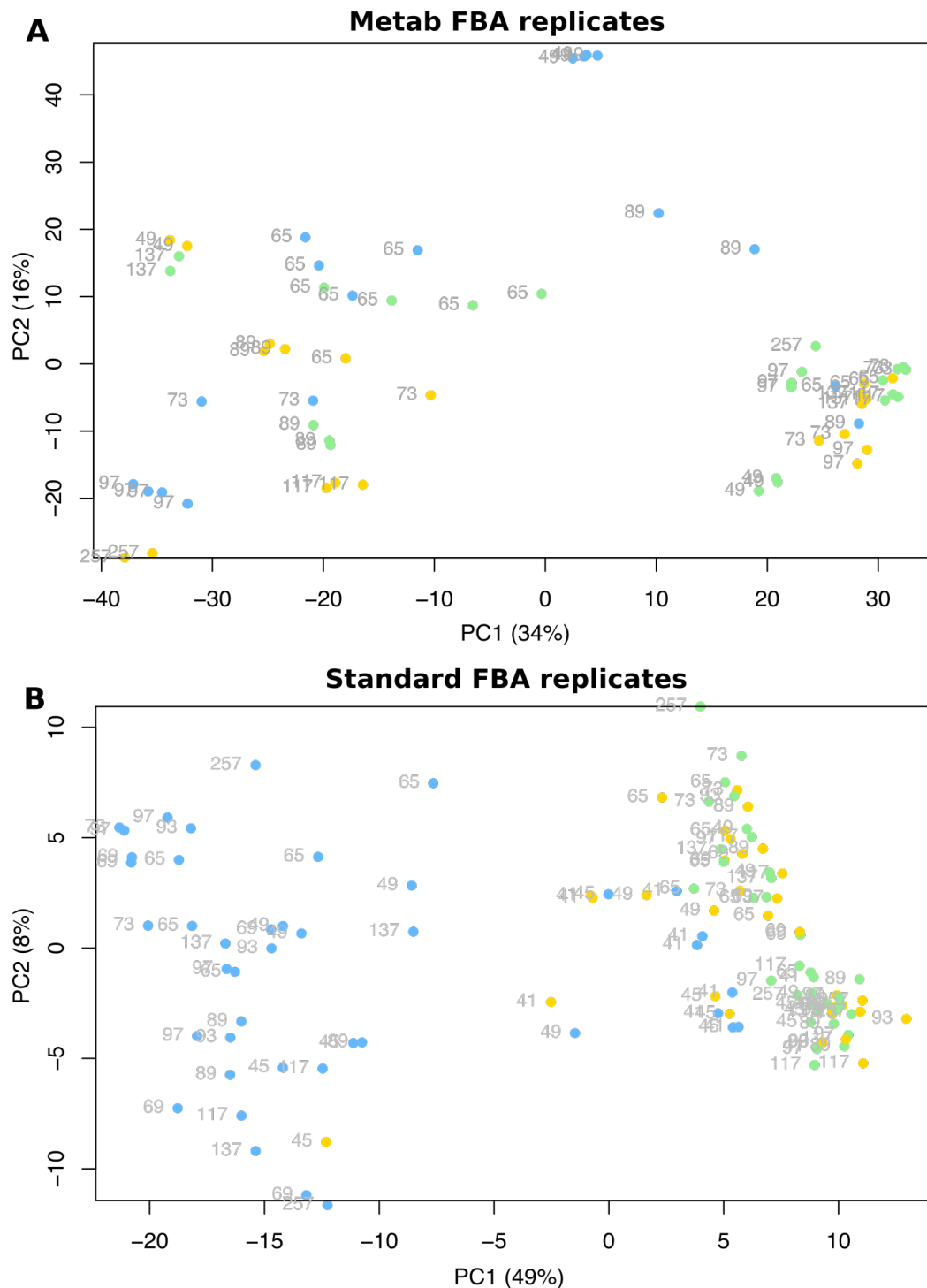


Figure A.6 PCA of flux values predicted for individual replicate samples.

In the FBA pipeline, flux values predicted using transcriptomics only (standard FBA), or transcriptomics with metabolomics (Metab FBA). To arrive at single flux predictions per group, the means of the transcriptomics replicates are used. However, to determine which reaction flux differences are significant, a distribution of flux values is needed, thus, flux predictions are made using the individual transcriptomics replicate values to give 'replicate' flux predictions. This Figure shows (A) the replicate flux predictions for the Metab FBA method, and (B) the replicate flux predictions for the standard FBA method. Colours indicate strain: blue for GLP, green for FEM and gold for GEM. Points are labelled with their collection time in hours. The Figure illustrates that replicates cluster together very well in Metab FBA, and broadly although not so well in standard FBA. However, this should be interpreted in the context of the fact that standard FBA flux values are less variable overall.

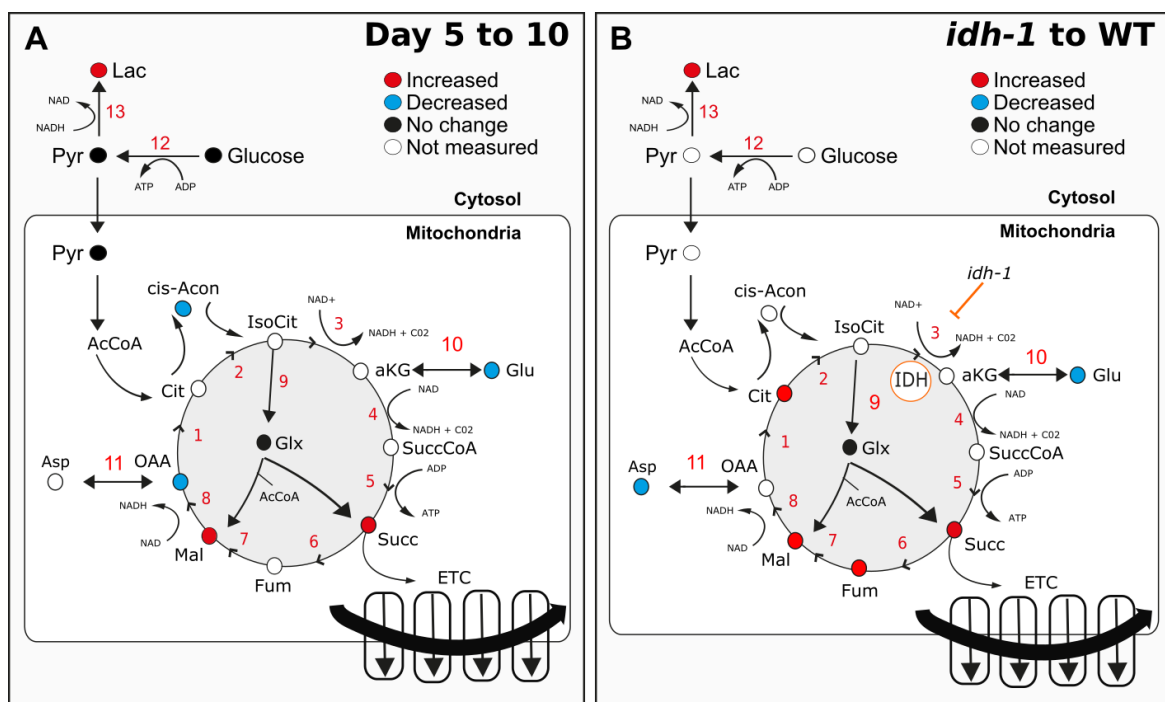


Figure A.7 Comparison of TCA cycle in aged worms to *idh-1* mutants.

Metabolite measurements show that aged worms have a defective TCA cycle. Metabolites that are found changed due to altered fluxes in (A) our dataset of old worms, and (B) *idh-1* mutants compared to wild type from (Vergano et al., 2014). The metabolites are color coded as follows: blue and red were found to be decreased and increased respectively; black represents metabolites that do not change with age and white represents unmeasured metabolites. This Figure first appeared in (Hastings et al., 2019a) and was originally created by Olivia Casanueva.

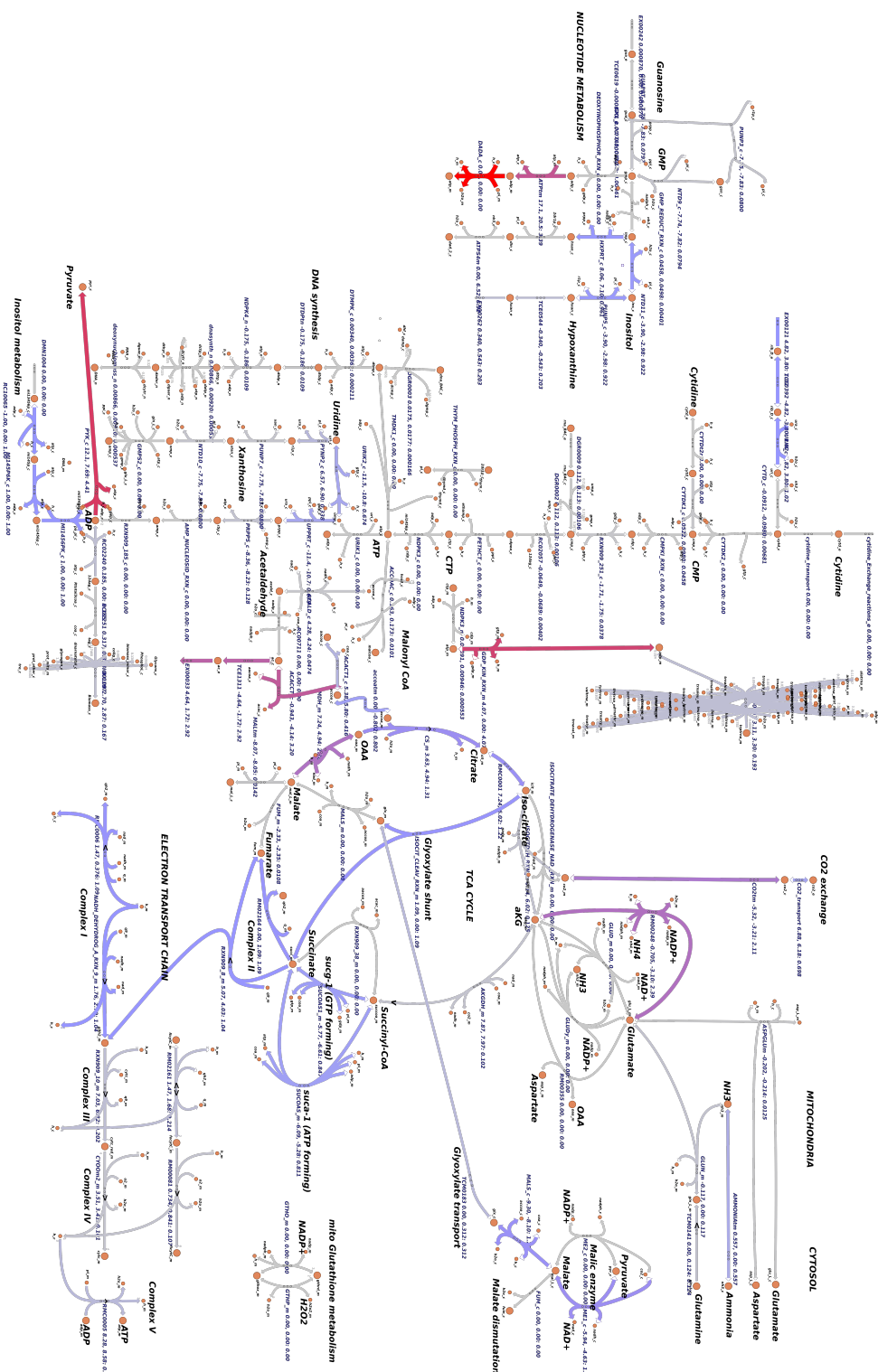


Figure A.8 Comparison of standard FBA fluxes between FEM and GLP at 49 h

The comparison of standard FBA fluxes for FEM and GLP at 49 h (Day 1) through an interconnected network of pathways including the TCA cycle and amino acid and glutamate metabolism is visualised in the Escher tool. Grey arrows indicate no difference in flux, blue indicates small differences only, while pink or red indicates large differences in compared fluxes. The colours do not provide information about whether the fluxes were increased or decreased in GLP relative to FEM, just the absolute magnitude of their difference. Reactions are labelled with their IDs as per the WormJam model as well as the predicted flux in each strain and the computed difference.

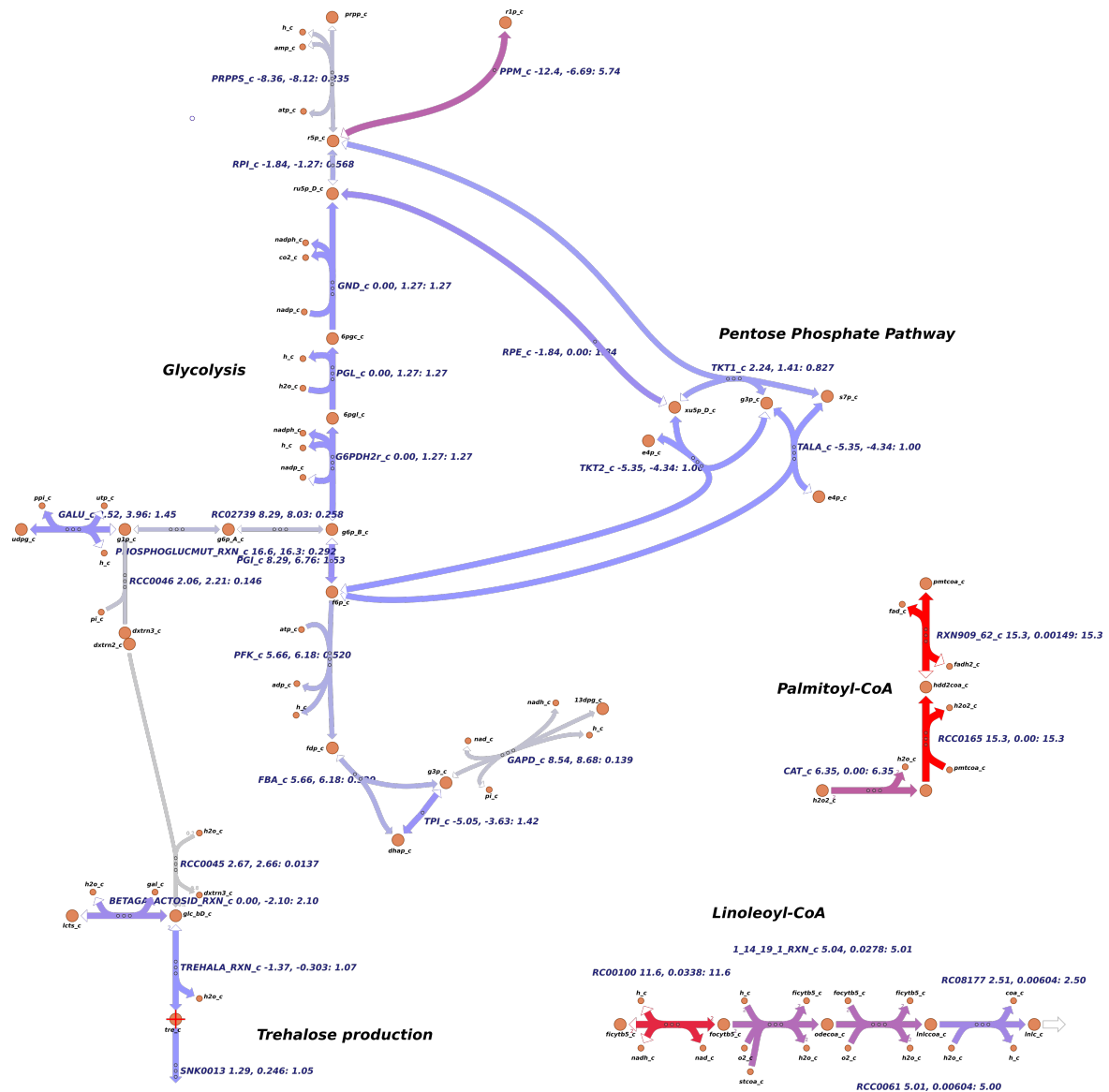


Figure A.9 Comparison of Metab FBA fluxes for several pathways between GLP 49 and 65 h

The comparison of metab FBA fluxes GLP between 49 h and 65 h (Day 1 and 2) through an interconnected network of pathways including the TCA cycle and glycolysis is visualised in the Escher tool. Grey arrows indicate no difference in flux, blue indicates small differences only, while pink or red indicates large differences in compared fluxes. The colours do not provide information about whether the fluxes were increased or decreased in GLP relative to FEM, just the absolute magnitude of their difference. Reactions are labelled with their IDs from the WormJam model, as well as the predicted flux at the two time points and the calculated difference.

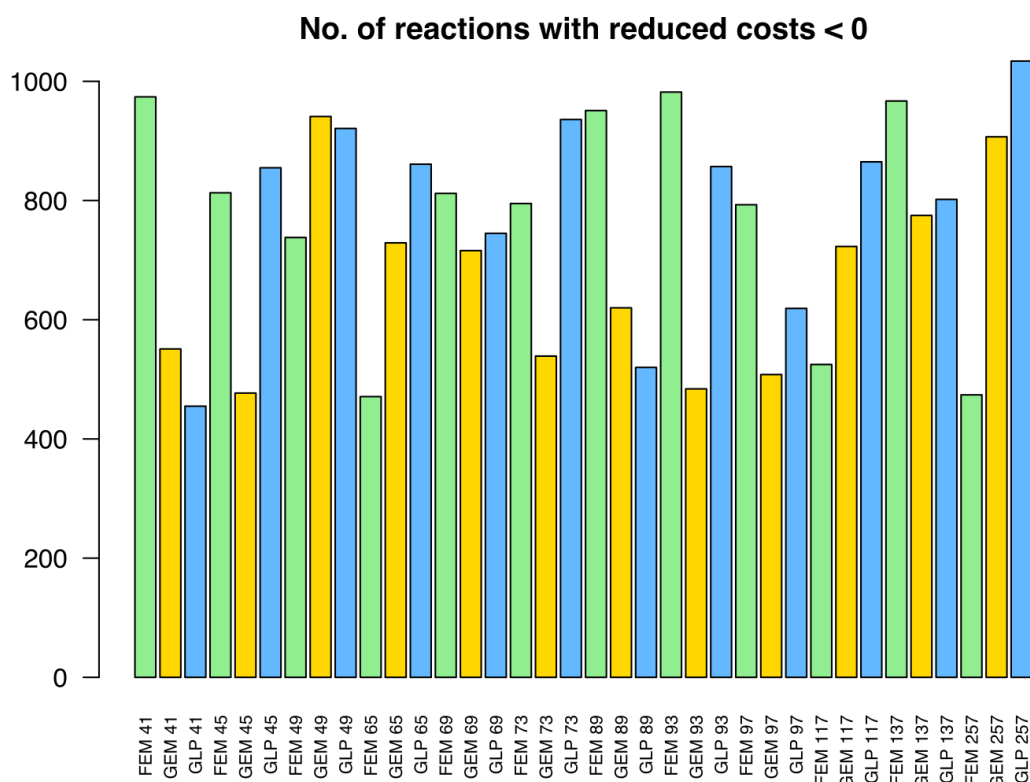


Figure A.10 Numbers of reactions with reduced costs < 0 per strain and time point

The Figure shows the number of reactions which have reduced costs < 0 per strain and time point calculated from the standard FBA model. There is no overall trend in any of the strains or with respect to age, however, it is clear that GLP has more such constrained reactions at several of the time points than the other two strains, indicating that GLP depends on more pathways acting at optimal level than the other strains. Colour indicates strain: blue for GLP, green for FEM and gold for GEM.